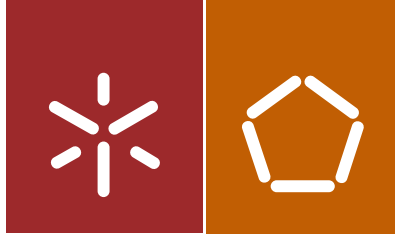


Universidade do Minho
Escola de Engenharia

João José Lopes Peixoto

Da Observação à Trajetória: formalização
de uma estrutura de informação
espaçotemporal

João José Lopes Peixoto Da Observação à Trajetória: formalização
de uma estrutura de informação espaçotemporal



Universidade do Minho
Escola de Engenharia

João José Lopes Peixoto

Da Observação à Trajetória: formalização
de uma estrutura de informação
espaçotemporal

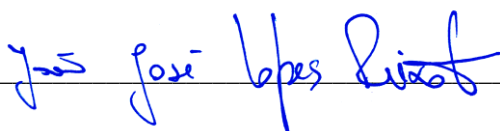
Tese de Doutoramento
Programa Doutoral em Tecnologias e Sistemas de
Informação

Trabalho efectuado sob a orientação do
Professor Doutor Adriano Moreira
Professor Doutor Carlos Bento

julho de 2017

É autorizada a reprodução parcial desta tese, apenas para efeitos de investigação, mediante declaração escrita do interessado, que a tal se compromete.

Universidade do Minho, 28/ 07/ 2017

Assinatura: _____

Declaração de Integridade

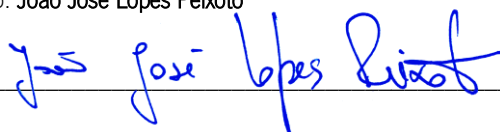
Declaro ter atuado com integridade na elaboração da presente tese. Confirmando que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Código de Conduta Ética da Universidade do Minho.

Universidade do Minho, 28/ 07 / 2017

Nome completo: João José Lopes Peixoto

Assinatura: _____



Agradecimentos

A conclusão deste trabalho exigiu um grande esforço e compreensão de pessoas que como eu não desistiram ao longo deste infindável período. Por esse motivo o meu grande agradecimento.

Aos meus familiares e amigos, que quando necessário estiveram sempre presentes, mesmo nas alturas em que o caminho mais fácil era desistir.

E por falar em desistir, um agradecimento particular ao meu orientador Professor Doutor Adriano Moreira, que nunca desistiu e teve a paciência de me apoiar durante esta longa caminhada.

A todos eles, o meu Muito Obrigado!

Apoio Financeiro

Este trabalho foi desenvolvido com o apoio da FCT e do programa POPH/FSE.



Resumo

A mobilidade dos cidadãos ao longo do tempo e espaço está na origem de múltiplos problemas que afetam o espaço urbano e quem nele habita. Congestionamentos de tráfego constantes, poluição ambiental nas grandes cidades, sistemas de transporte sobrelotados ou desadequados, são alguns desses problemas.

Recentemente têm sido muitos os trabalhos focados na análise de padrões de mobilidade com o objetivo de ajudar na resolução destes problemas. Com recurso a diferentes tipos de sensores, estes trabalhos pretendem estudar vários aspetos relativos às dinâmicas urbanas.

O recurso a diferentes tipologias de sensores e consequentemente diferentes estruturas de dados tem dificultado a estruturação de conceitos e o uso da informação na representação da mobilidade humana.

No âmbito deste trabalho foi desenvolvida uma estrutura de conceitos de mobilidade e processos de transformação de modo a acomodar dados espaçotemporais obtidos a partir de diferentes sensores, cujos dados são muito diferentes entre si. Esta estrutura de conceitos foi posteriormente validada com duas experimentações distintas.

A primeira experimentação teve por base o desenvolvimento de um modelo probabilístico cujo objetivo principal visa permitir a fusão de dados de múltiplos sensores e a partir destes, a extração dos diferentes conceitos definidos na estrutura desenvolvida.

O desenvolvimento da segunda experimentação focou a utilização de dados de redes sociais e a análise sobre o potencial dos mesmos na sua utilização em análises de mobilidade humana.

As duas experimentações, realizadas para diferentes fontes de dados, permitiram a validação da integração de dados geométricos e simbólicos associados à informação de posicionamento dos artefactos.

Uma visão do que poderia ser um Ecossistema Urbano de apoio ao planeamento e decisão com base na análise de mobilidade humana é também apresentado como forma de incentivo à colaboração dos decisores e analistas na resolução de problemas e anomalias existentes nos espaços urbanos. Nesta visão é dado particular destaque à criação de mapas individuais e globais de mobilidade, quer nas vantagens e desvantagens das duas abordagens, quer nos resultados que se conseguem obter. A inclusão no mesmo estudo de mapas obtidos através de dados agregados e de observações individuais de artefactos permite enriquecer a informação que está em análise potenciando a extração de conclusões mais ricas, mais próximas da realidade e mais abrangentes.

Abstract

The mobility of citizens over the time and space is the source of many problems that affect the urban space and those living in it. Recurrent traffic jams, environmental pollution in large cities, overcrowded or inadequate transport systems, are some of these problems.

Recently, there have been many studies focused on the analysis of mobility patterns in order to help solve these problems. Using different types of sensors, these works intend to study various aspects relating to urban dynamics.

The use of different types of sensors and consequently different data structures have hampered the structuring of concepts and the usage of the information in the representation of human mobility.

In this work it was developed a framework of mobility concepts and transformation processes, to accommodate spatiotemporal data from different sensors, whose data are very different. This structure of concepts was subsequently validated using two distinct experiments.

The first experiment was based on the development of a probabilistic model whose main purpose is to allow the data fusion of multiple sensor and from these, the extraction of different concepts associated to the developed framework.

The development of second experiment was focused in the use of social networking data and the analysis of the potential of their use in human mobility analysis.

The two experiments, conducted for different sources of data, allowed the validation of the integration of geometric and symbolic data associated with the position information of the artefacts.

A vision of what could be an Urban Ecosystem to support decision and urban planning based on the human mobility analysis is also presented in this work, as a way to encourage the collaboration between decision-makers and analysts in solving problems and anomalies that exist in urban areas. In this vision, particular emphasis is given to the creation of individual and global mobility maps,

highlighting the advantages and disadvantages of both approaches, as well as the results that can be achieved. The inclusion in the same study of mobility maps obtained through aggregated data and individual observations of artefacts allows to enrich the information that is being analysed, potentiating the extraction of richer conclusions, closer to reality and more comprehensive.

Índice

DECLARAÇÃO DE INTEGRIDADE	V
AGRADECIMENTOS.....	VII
APOIO FINANCEIRO	VII
RESUMO.....	IX
ABSTRACT	XI
ÍNDICE	XIII
LISTA DE ACRÓNIMOS	XVII
GLOSSÁRIO.....	XIX
ÍNDICE DE FIGURAS.....	XXIII
ÍNDICE DE TABELAS	XXVII
CAPÍTULO 1. INTRODUÇÃO E OBJETIVOS	1
1.1. ENQUADRAMENTO.....	1
1.2. OBJETIVOS E CONTRIBUIÇÕES DA TESE	3
1.3. ORGANIZAÇÃO DO DOCUMENTO	5
1.4. CONTRIBUTOS ORIGINAIS.....	6
CAPÍTULO 2. PERSPETIVAS SOBRE A MOBILIDADE HUMANA	7
2.1. ÁREAS APLICACIONAIS.....	8
2.2. NATUREZA DO MOVIMENTO HUMANO.....	11
2.3. VISUALIZAÇÃO DO MOVIMENTO HUMANO	17
2.4. OBSERVAÇÃO DO MOVIMENTO HUMANO	36
2.5. TÉCNICAS DE ANÁLISE	52
2.6. DESAFIOS AINDA SEM RESPOSTA	56

CAPÍTULO 3. ECOSSISTEMA URBANO.....	59
3.1. OBSERVAÇÃO E ANÁLISE.....	60
3.2. ATUAÇÃO	64
3.3. ECOSSISTEMA DE MOBILIDADE URBANA.....	66
CAPÍTULO 4. ESTRUTURA DE CONCEITOS	71
4.1. ESTRUTURA DE CONCEITOS.....	71
4.1.1. <i>Dados em bruto</i>	72
4.1.2. <i>Observação</i>	74
4.1.3. <i>Lugar</i>	76
4.1.4. <i>Suspensão do movimento</i>	80
4.1.5. <i>Movimento</i>	81
4.1.6. <i>Trajectoria</i>	84
4.1.7. <i>Processos de transformação</i>	86
CAPÍTULO 5. EXPERIMENTAÇÃO	89
5.1. INTRODUÇÃO.....	89
5.2. DADOS INDIVIDUAIS – MÚLTIPLOS SENSORES	89
5.2.1. <i>Aquisição dos Dados</i>	90
5.2.2. <i>Dos Dados em bruto às Observações</i>	95
5.2.3. <i>Lugares e Estadias</i>	96
5.2.4. <i>Suspensão do movimento</i>	112
5.2.5. <i>Movimento</i>	114
5.2.6. <i>Trajectoria</i>	121
5.3. REDES SOCIAIS	123
5.3.1. <i>Aquisição dos Dados</i>	123
5.3.2. <i>Dos Dados em bruto às Observações</i>	127
5.3.3. <i>Lugares e Estadias</i>	129
5.3.4. <i>Suspensão do movimento</i>	139
5.3.5. <i>Movimento</i>	140
5.3.6. <i>Trajectoria</i>	144
5.4. CONCLUSÕES.....	145
CAPÍTULO 6. CONCLUSÕES.....	155
6.1. ECOSSISTEMA URBANO	155
6.2. ESTRUTURA DE CONCEITOS.....	156
6.3. FUSÃO DE DADOS.....	157
6.4. DADOS EM BRUTO	158

6.5. RESULTADOS DAS IMPLEMENTAÇÕES DOS CONCEITOS	159
6.6. FUTURAS EVOLUÇÕES	160
BIBLIOGRAFIA.....	163

Lista de acrónimos

AP	Access Point
API	Application Programming Interface
BSSID	Basic Service Set Identification
CID	Cell Id
GIS	Geographic Information System
GPS	Global Positioning System
GSM	Global System for Mobile Communications
JSON	JavaScript Object Notation
LAC	Location Area Code
MAC	Media Access Control
MIT	Massachusetts Institute of Technology
MNC	Mobile Network Codes
RSSI	Received Signal Strength Indication
SSID	Service Set Identifier
Wi-Fi	Wireless local area networking

Glossário

B

BACKGROUND – termo aplicado a aplicações em execução em segundo plano, sem necessidade de intervenção do utilizador.

BIKE SHARING – sistemas de partilha de bicicletas.

C

CARPOOLING – solução de partilha de carro com outros passageiros com destinos e horários coincidentes. O proprietário da viatura partilha total ou parcialmente o seu trajeto com outros passageiros.

CHECK-IN – ato de assinalar a entrada/chegada a lugar específico.

CHECK-OUT – ato de assinalar a saída de um lugar específico.

CLUSTER – aglomerado de coisas semelhantes.

CLUSTERING – algoritmo usado para agregação de coisas semelhantes.

D

DATASET – conjunto de dados.

E

ELEMENTARY MOVEMENT – conceito que caracteriza movimentos realizados num intervalo de tempo de tal forma reduzido que é possível extrapolar a trajetória com exatidão.

F

FINGERPRINT – no contexto deste trabalho é uma amostra composta por dados de um ou mais sensores, onde a estrutura de dados de cada sensor é regra geral muito própria e com reduzidas semelhanças entres sensores, e que caracterizam de forma única um determinado local.

FOOTPRINT – pegada eletrónica deixada ao longo do tempo pelos utilizadores.

H

HEATMAP – representação gráfica com base em classificação por cores (cores quentes e cores frias), com aplicação para a representação da ocupação do espaço urbano.

L

LOCATION – no contexto deste trabalho, define uma posição/local num determinado referencial geográfico ou geométrico.

M

MERGE – ação de agregar, neste caso em particular, agregar dois ou mais locais num único local.

O

OBSERVATION – no contexto deste trabalho, caracteriza o momento e o local onde um objeto ou pessoa foi observado do ponto de vista da mobilidade.

OPEN DATA – conceito aplicado a dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa.

OUTLIERS – termo utilizado frequentemente nos algoritmos de *clustering*, para definir um valor discrepante do(s) aglomerados que foram identificados.

OVERFITTING – termo utilizado para classificar um modelo que se ajusta demasiado ao conjunto de dados usados.

P

PARSER – interpretador e/ou conversor de dados.

PLACE – lugar ou área geográfica identificado como relevante ou candidato de acordo com o efetivo significado para o utilizador, onde este permaneceu por um período de tempo.

PLACE LEARNING – algoritmos desenvolvidos para a extração automática de *Places*.

PROXY – no contexto deste trabalho refere-se a objetos ou aplicações que por acompanharem as pessoas no seu dia-a-dia refletem através da sua observação uma realidade ao nível da mobilidade muito semelhante à mobilidade desses indivíduos.

Q

QUERY – comando SQL que permite a consulta, adição e alteração de dados e tabelas num motor de base de dados.

R

RANDOM WALK MODEL – modelo que descreve um trajeto com base numa sucessão de movimentos aleatórios.

RAW DATA – dados em bruto obtidos diretamente dos sensores, sem qualquer processamento.

S

SNAPSHOT – registo instantâneo, que no âmbito deste trabalho, refere-se a uma foto/imagem que caracteriza naquele momento o espaço urbano.

SPACE LEAP – conceito que define um salto no referencial geométrico/geográfico, que por falta de observação da pessoa/objeto durante esse período, não permite com exatidão extrapolar se houve ou não movimentação.

SPLIT – ação de dividir, neste caso em particular, dividir um local em vários.

SQL – linguagem estruturada para gestão de dados dentro de um motor de base de dados.

STAY – conceito que define o período de tempo em que uma pessoa/objeto permaneceu num determinado *Place*, não havendo por isso evidências de movimentação.

T

TIMESTAMP – instante temporal (data e hora) onde um objeto/pessoa foi observado.

TIME LEAP – conceito que define um período temporal alargado durante o qual não existem evidências de que tenha havido qualquer movimento do objeto/pessoa, por não existir informação relativa à sua observação.

TRACES (GPS) – rasto de movimentos obtidos, neste caso em particular, através dos recetores GPS.

TRAJECTORY – conceito que no âmbito deste trabalho representa uma lista ordenada de movimentações de uma pessoa/objeto ao longo do tempo.

TRUNCATED POWER LAW – distribuição estatística com base na lei de potências.

TWEET – publicação realizada na rede social *Twitter*.

V

VISUAL ANALYTICS – processo de análise visual com base na manipulação dos dados de forma a realizarmos um conjunto de validações ou testes de hipóteses.

W

WIRELESS – rede sem fios.

Índice de figuras

FIGURA 2.1 - EXEMPLOS DE ALGUMAS ÁREAS APLICACIONAIS	11
FIGURA 2.2 - FREQUÊNCIA DAS VISITAS AOS LOCAIS (GONZÁLEZ, HIDALGO E BARABÁSI, 2008)	13
FIGURA 2.3 – RELAÇÃO ENTRE OS 10 LOCAIS MAIS VISITADOS E A SUA PREVISIBILIDADE (SONG ET AL. 2010)	14
FIGURA 2.4 – DISTRIBUIÇÃO DAS DISTÂNCIAS (BROCKMANN, HUFNAGEL E GEISEL, 2006)	16
FIGURA 2.5 – VISUALIZAÇÃO DA OCUPAÇÃO DO ESPAÇO AO LONGO DO TEMPO (READES ET AL., 2007)	17
FIGURA 2.6 – VISUALIZAÇÃO DA CIDADE DE ROMA EM DIFERENTES HORAS DO DIA (READES ET AL., 2007)	18
FIGURA 2.7 - VISUALIZAÇÃO DE ZONAS COM MAIOR CONCENTRAÇÃO DE PESSOAS (G. ANDRIENKO, ANDRIENKO, BOSCH, ET AL., 2013) ...	19
FIGURA 2.8 – DENSIDADE TEMPORAL (JIANG, FERREIRA E GONZALEZ, 2012)	20
FIGURA 2.9 – VISUALIZAÇÃO DE VELOCIDADES MÉDIAS AO LONGO DO TEMPO E NO ESPAÇO (ANDRIENKO ET AL., 2010)	21
FIGURA 2.10- TRAJETÓRIAS DE VEÍCULOS (A) E ZONAS DE VELOCIDADE BAIXA (B) (ANDRIENKO ET AL., 2011)	22
FIGURA 2.11 – TRAJETÓRIAS DE VOOS (ANDRIENKO ET AL., 2011)	23
FIGURA 2.12 – REPRESENTAÇÃO DAS TRAJETÓRIAS DE NOTAS (BROCKMANN E THEIS, 2008)	24
FIGURA 2.13 – REPRESENTAÇÃO BASEADA EM GRAFOS (SONG ET AL., 2010)	25
FIGURA 2.14 – REPRESENTAÇÃO DE LOCAIS VISITADOS E TRAJETÓRIAS (G. ANDRIENKO, ANDRIENKO, FUCHS, ET AL., 2013)	26
FIGURA 2.15 – REPRESENTAÇÃO ESPAÇOTEMPORAL DE TRAJETÓRIAS E LOCAIS (G. ANDRIENKO, ANDRIENKO, FUCHS, ET AL., 2013)	27
FIGURA 2.16 – CÓDIGO EXEMPLO DE UM PROJETO EM PROCESSING	28
FIGURA 2.17 – VOOS COMERCIAIS REALIZADOS NUM DIA NOS EUA	29
FIGURA 2.18 – CONGESTIONAMENTO NA CIDADE DE ILLINOIS (PLANNING, 2014)	31
FIGURA 2.19 – SAÍDA PARA O TRABALHO, CASO 1 (YAU, 2014)	32
FIGURA 2.20 – SAÍDA PARA O TRABALHO, CASO 2 (YAU, 2014)	33
FIGURA 2.21 – PROJETO MANYCITIES (GRAUWIN ET AL., 2014)	33
FIGURA 2.22 – EXEMPLO DA SELEÇÃO DOS OBJETOS DE ANÁLISE (CABALERO E BELMONTE, 2013)	34
FIGURA 2.23 – MIGRAÇÃO AO LONGO DOS ESTADOS UNIDOS DA AMÉRICA (BRUNER, 2012)	34
FIGURA 2.24 – FLUXO DE TÁXIS AO LONGO DO TEMPO (BLOCH, FESSENDEN E CARTER, 2009)	35
FIGURA 2.25 – TAXA DE UTILIZAÇÃO DO METRO AO LONGO DO TEMPO (ESRI, 2016)	35
FIGURA 2.26 – ASSINATURA MAGNÉTICA APÓS TRÊS PASSAGENS DE VEÍCULOS (AHDİ ET AL., 2012)	46

FIGURA 2.27 – DETEÇÃO AUTOMÁTICA DE VEÍCULOS POR INFRAVERMELHOS (BACON ET AL., 2011)	47
FIGURA 2.28 – EXEMPLO DA IDENTIFICAÇÃO DE VEÍCULOS COM RECURSO A IMAGEM VÍDEO (KOLLER, DANILIDIS E NAGEL, 1993)	47
FIGURA 2.29 – TIPOLOGIAS DE OBSERVAÇÕES.....	52
FIGURA 3.1 - OBSERVAÇÃO E SENSORES.....	59
FIGURA 3.2 – EXEMPLOS DO PROCESSO DE CRIAÇÃO DE MAPAS DE MOBILIDADE.....	60
FIGURA 3.3 – REPRESENTAÇÃO DOS RESULTADOS DA ANÁLISE DE MOBILIDADE.....	61
FIGURA 3.4 - PROPOSTA PARA A ELABORAÇÃO DE ESTUDOS DE MOBILIDADE.....	63
FIGURA 3.5 – ECOSSISTEMA DE MOBILIDADE URBANA	67
FIGURA 4.1 – OBSERVAÇÕES NUM REFERENCIAL ESPAÇOTEMPORAL (A) E A MULTIPLICIDADE DAS DIMENSÕES ESPACIAIS (B)	75
FIGURA 4.2 – EXTRAÇÃO DE UM LUGAR	77
FIGURA 4.3 – UM PLACE COM QUARTO OBSERVATIONS DIFERENTES E UM TIMELEAP ENTRE DUAS DELAS	81
FIGURA 4.4 – SEQUÊNCIA DE ELEMENTARY MOVEMENTS ENTRE LOCATIONS	83
FIGURA 4.5 – UM EXEMPLO DE UM SPACE LEAP.....	84
FIGURA 4.6 – O CONCEITO DE TRAJECTORY.....	85
FIGURA 4.7 – ESTRUTURA EM CAMADAS DOS CONCEITOS DE INFORMAÇÃO	86
FIGURA 4.8 – CONCEITO PLACE DETALHADO	87
FIGURA 5.1 – COMPARATIVO ENTRE O TEMPO IMÓVEL E EM VIAGEM.....	91
FIGURA 5.2 – PERCENTAGEM DE TEMPO NOS 5 LOCAIS MAIS RELEVANTES E RESTANTES LOCAIS	92
FIGURA 5.3 – DISTRIBUIÇÃO DOS DADOS DE ACORDO COM A TIPOLOGIA.....	94
FIGURA 5.4 – UM PLACE CARACTERIZADO POR DIFERENTES TIPOS DE OBSERVAÇÕES	100
FIGURA 5.5 – ALGORITMO DA EXTRAÇÃO DO PLACE.....	101
FIGURA 5.6 – PLACES CANDIDATOS DE UM INDIVÍDUO.....	102
FIGURA 5.7 – ALGORITMO DA EXTRAÇÃO DO STAYS	104
FIGURA 5.8 – STAYS NO ESPAÇO E TEMPO.....	104
FIGURA 5.9 – AGREGAÇÃO DE STAYS – EXEMPLO 1	105
FIGURA 5.10 – AGREGAÇÃO DE STAYS – EXEMPLO 2	106
FIGURA 5.11 – ALGORITMO DA ELEVAÇÃO DO PLACE DE CANDIDATO A RELEVANTE	107
FIGURA 5.12 – PLACES CANDIDATOS (ESQUERDA) E PLACES RELEVANTES (DIREITA)	108
FIGURA 5.13 – LOCALIZAÇÃO DE UM PLACE EXTRAÍDO	111
FIGURA 5.14 – PLACE IDENTIFICADO NUMA VIA DE COMUNICAÇÃO.....	112
FIGURA 5.15 – OBSERVAÇÕES, STAYS E TIME LEAP	112
FIGURA 5.16 – PLACES RELEVANTES.....	115
FIGURA 5.17 – ESTADOS DE MOBILIDADE DOS ARTEFACTOS.....	115
FIGURA 5.18 – EXEMPLO DE ELEMENTARY MOVEMENTS	117
FIGURA 5.19 – DETALHE COM VÁRIOS ELEMENTARY MOVEMENTS.....	118
FIGURA 5.20 – EXEMPLOS DE SPACE LEAPS.....	119
FIGURA 5.21 – EXEMPLO DA DIFICULDADE NA REPRESENTAÇÃO NO GEPHI	119
FIGURA 5.22 – SPACE LEAPS APENAS SIMBÓLICOS	121

FIGURA 5.23 – EXEMPLO DE TRAJECTORY	122
FIGURA 5.24 – PLACES, ELEMENTARY MOVEMENTS E SPACE LEAPS	122
FIGURA 5.25 – ESTRUTURA DO OBJECTO JSON.....	125
FIGURA 5.26 – CHECKIN NUM RESTAURANTE	130
FIGURA 5.27 – TWEETS NO TEMPO.....	131
FIGURA 5.28 – HEATMAP COM AS LOCALIZAÇÕES DOS TWEETS DE UM UTILIZADOR	132
FIGURA 5.29 – LOCAIS COM MAIOR OCORRÊNCIA (A), LOCAIS COINCIDENTES COM VIAS (B).....	133
FIGURA 5.30 - OCUPAÇÃO DE ESPAÇO OBTIDO ATRAVÉS DE 2340 TWEETS REALIZADOS POR 867 UTILIZADORES	134
FIGURA 5.31 – CLASSIFICAÇÃO DE OCORRÊNCIAS POR CORES	134
FIGURA 5.32 – OUTRA REPRESENTAÇÃO DA OCUPAÇÃO DE ESPAÇO COM DIFERENTES PARÂMETROS DE CLASSIFICAÇÃO	135
FIGURA 5.33 – DETALHE DE UMA ZONA GEOGRÁFICA.....	135
FIGURA 5.34 – EVOLUÇÃO DA OCUPAÇÃO DE ESPAÇO.....	137
FIGURA 5.35 – REPRESENTAÇÃO GRÁFICA DOS SPACE LEAPS	142
FIGURA 5.36 – SPACE LEAPS DE OBSERVAÇÕES TWITTER.....	142
FIGURA 5.37 – LOCALIZAÇÃO DOS PLACES	143
FIGURA 5.38 – REPRESENTAÇÃO DO CONCEITO TRAJECTORY EM DOIS DIAS DISTINTOS PARA O MESMO UTILIZADOR.....	144
FIGURA 5.39 – ESTRUTURA DE CONCEITOS AJUSTADA ÀS REDES SOCIAIS.....	148
FIGURA 5.40 – QUATRO PERFIS DE MOBILIDADE DE QUATRO UTILIZADORES PARA O MESMO REFERENCIAL ESPAÇOTEMPORAL	150
FIGURA 5.41 – MAPA GLOBAL APÓS AGREGAÇÃO DOS MAPAS INDIVIDUAIS DE MOBILIDADE.....	151
FIGURA 5.42 – LOCAIS COM MAIOR NÚMERO DE PUBLICAÇÕES.....	152

Índice de tabelas

TABELA 2.1 – INFORMAÇÃO MÍNIMA EXTRAÍDA DE UM TWEET	38
TABELA 2.2 – POTENCIAL INFORMAÇÃO OBTIDA DAS REDES CELULARES.....	40
TABELA 2.3 – OUTRA POTENCIAL REPRESENTAÇÃO DOS DADOS DAS REDES CELULARES	40
TABELA 2.4 – INFORMAÇÃO PARCIAL RECOLHIDA SOBRE A UTILIZAÇÃO DA REDE Wi-Fi	41
TABELA 2.5 – DADOS DE BILHÉTICA (LIU, BIDERMAN E RATTI, 2009)	41
TABELA 2.6 – ESTRUTURA DE DADOS GPS.....	42
TABELA 2.7 – DADOS OBTIDOS DA INFRAESTRUTURA DE BIKE SHARING.....	45
TABELA 2.8 – EXEMPLO DOS DADOS Wi-Fi RECOLHIDOS	51
TABELA 2.9 – EXEMPLO DE DADOS DA REDE CELULAR	51
TABELA 4.1 – DADOS EM BRUTO DO SENSOR GPS	72
TABELA 4.2 – DADOS EM BRUTO DO SENSOR Wi-Fi.....	72
TABELA 4.3 – DADOS EM BRUTO DO SENSOR GSM.....	72
TABELA 5.1 – EXEMPLO DOS DADOS GPS RECOLHIDOS	93
TABELA 5.2 – EXEMPLO DOS DADOS Wi-Fi RECOLHIDOS	93
TABELA 5.3 – EXEMPLO DOS DADOS GSM RECOLHIDOS	94
TABELA 5.4 – ESTRUTURA DE DADOS DO CONCEITO OBSERVATION.....	95
TABELA 5.5 – EXEMPLO DO RESULTADO DO PRIMEIRO PROCESSO DE TRANSFORMAÇÃO.....	96
TABELA 5.6 – MATRIZ DE PROBABILIDADES ENTRE DUAS OBSERVAÇÕES.....	98
TABELA 5.7 – VALORES DOS PARÂMETROS UTILIZADOS NAS PROBABILIDADES	99
TABELA 5.8 – ESTRUTURA DE DADOS DO CONCEITO PLACE.....	100
TABELA 5.9 – ESTRUTURA IMPLEMENTADA PARA O CONCEITO PLACE	100
TABELA 5.10 – RESULTADO DA FILTRAGEM E AGREGAÇÃO DE STAYS	106
TABELA 5.11 – EXEMPLO DA ESTRUTURA DE UMA STAY	106
TABELA 5.12 – RESULTADO DO TERCEIRO PROCESSO DE TRANSFORMAÇÃO	107
TABELA 5.13 – COMPARATIVO ENTRE OS PLACES DO DIÁRIO E OS PLACES EXTRAÍDOS PELO ALGORITMO.....	109
TABELA 5.14 – DIÁRIO ANOTADO COM ADIÇÃO DOS PLACES EXTRAÍDOS	109
TABELA 5.15 – EXTRACTO DE STAYS POTENCIAIS.....	110

TABELA 5.16 – ESTRUTURA DO CONCEITO TIME LEAP	113
TABELA 5.17 – EXEMPLO DA REPRESENTAÇÃO DOS ELEMENTARY MOVEMENTS.....	116
TABELA 5.18 – EXEMPLO DA REPRESENTAÇÃO DOS SPACE LEAPS.....	117
TABELA 5.19 – RAW DATA IMPORTADOS	126
TABELA 5.20 – RAW DATA CHECKINS	127
TABELA 5.21 – RAW DATA VENUES	127
TABELA 5.22 – EXEMPLO DE OBSERVATIONS DE REDES SOCIAIS	128
TABELA 5.23 – ESTRUTURA DE DADOS DO CONCEITO PLACE.....	136
TABELA 5.24 – PERCENTAGENS DE CHECKINS NUM SÁBADO.....	138
TABELA 5.25 – EXEMPLOS DE OBSERVAÇÕES NO MESMO PLACE	139
TABELA 5.26 – EXEMPLOS DE TIME LEAPS EXTRAÍDOS	140
TABELA 5.27 – EXEMPLO DE TRÊS SPACE LEAPS DE UM UTILIZADOR	141
TABELA 5.28 – RESULTADOS DA FUSÃO DE DADOS	146

Capítulo 1. Introdução e objetivos

1.1. Enquadramento

Nos últimos anos, conceitos e trabalhos realizados na área da mobilidade urbana e cidades inteligentes foram alvo de grande evolução. A maior consciencialização para os problemas existentes nas nossas cidades e a forma como estes influenciam a qualidade de vida das pessoas levou a que este tipo de problemática fosse cada vez mais explorado.

O custo na criação ou modificação de infraestruturas dentro de um espaço urbano é outro dos motivos de interesse para o estudo da mobilidade urbana. Qualquer infraestrutura urbana tem normalmente um grande impacto financeiro, pelo que o investimento apenas se justifica se efetivamente constituir uma mais-valia para os cidadãos e para o próprio espaço. Apenas a título de exemplo, em 2013 a cidade de Londres estimava gastar, até 2023, cerca de mil milhões de euros na rede de ciclovias da cidade (Authority, 2013). Atendendo à dimensão do investimento, seria imprudente que a infraestrutura fosse inadequada à realidade da mobilidade desta cidade. No entanto, na maioria das situações de planeamento de cidades e outros espaços urbanos não existe uma real noção das necessidades e comportamentos de mobilidade, sendo cada vez mais comum o aparecimento de aplicações isoladas ou plataformas mais elaboradas visando a compreensão da mobilidade urbana.

Uma das formas de tentar compreender o movimento humano passa pela análise estatística de dados provenientes de várias fontes, como o trabalho apresentado por González (González, Hidalgo e Barabási, 2008) que concluiu que as trajetórias humanas apresentam um padrão e, em particular, que por norma o ser humano passa cerca de metade do seu tempo em apenas dois locais e que os locais mais frequentados por cada indivíduo são na sua maioria apenas cinco. Estes resultados são interessantes não só pela caracterização que foi feita ao nível dos locais e trajetórias entre eles, mas

porque indiciam que o movimento humano é repetitivo sendo possível a sua previsão com base na análise do seu passado.

Esta hipótese tem motivado outros investigadores para a análise do movimento humano em espaços urbanos. Um dos primeiros trabalhos nessa área passou pela caracterização do movimento humano ao longo de diferentes dias para diferentes locais de uma cidade (Reades *et al.*, 2007). Com este trabalho, Reads demonstrou diferentes padrões de ocupação do espaço de acordo com a tipologia do local e padrões diferentes no espaço urbano para os dias úteis comparativamente com o fim de semana. Assim como Reads, também Andrienko é outro investigador com trabalho relevante na caracterização do movimento humano em espaços urbanos. Quer através de análise dos locais com maior concentração de indivíduos num espaço urbano (G. Andrienko, Andrienko, Bosch, *et al.*, 2013); através da representação de locais de congestionamento de tráfego com base na velocidade média de veículos equipados com recetor GPS (Andrienko *et al.*, 2010) e respetivas trajetórias (Andrienko *et al.*, 2011); ou ainda pela representação de locais relevantes para os indivíduos e trajetórias entre esses locais (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013).

Além dos autores referidos anteriormente, a observação das dinâmicas das cidades, e respetivas análises e visualizações, têm suscitado muito interesse na comunidade científica, existindo por isso diversos trabalhos com técnicas de observação e técnicas de representação e visualização diferentes.

A técnica mais antiga de observação da mobilidade humana passa pela realização de inquéritos ou questionários à população. Embora seja uma técnica com um custo significativo e os dados possam rapidamente ficar desatualizados, ainda existem trabalhos onde este mecanismo de recolha de dados, ou observação do espaço urbano, é aplicado (Jiang, Ferreira e Gonzalez, 2012).

No entanto, o avanço tecnológico tem permitido a utilização de outros mecanismos de sensorização do espaço urbano. Desde a utilização de dados de recetores GPS (Andrienko *et al.*, 2010), redes Wi-Fi (Sevtsuk e Ratti, 2005), redes celulares (Reades *et al.*, 2007), passando pelas redes sociais (MacEachren e Robinson, 2011), e terminando em dados de bilhética (Liu, Biderman e Ratti, 2009), são já várias as tecnologias usadas para a observação de artefactos e respetivas movimentações ao longo do espaço urbano.

1.2. Objetivos e contribuições da tese

Embora exista atualmente uma variedade de abordagens e trabalhos realizados com o objetivo de estudar a mobilidade urbana e suas implicações, a falta de conceitos e estratégias para a representação da mobilidade levou a que cada autor definisse a sua própria nomenclatura, muitas vezes ajustada apenas ao seu problema concreto.

A existência de diferentes técnicas de observação do espaço urbano suportadas por uma recolha de dados utilizando diferentes tecnologias origina diferentes representações de informação e dificulta a uniformização de conceitos associados à mobilidade humana.

O objetivo principal desta tese visa colmatar essa dificuldade na representação da informação de mobilidade, através da definição e desenvolvimento de uma estrutura de conceitos suficientemente capaz de incluir os conceitos básicos associados à descrição do que é a mobilidade humana, bem como representar informação espaçotemporal oriunda de múltiplos e diferentes sensores.

Como resultado deste trabalho, em particular com o desenvolvimento da estrutura de conceitos apresentada no capítulo 4, pretendemos contribuir para uma clarificação de conceitos e processos de transformação de dados/informação ao nível da representação da mobilidade urbana, algo que estava ainda muito disperso e sem uma estrutura clara nos vários trabalhos desenvolvidos na área.

Com a implementação da estrutura de conceitos e processos de transformação associados, descrita ao longo do capítulo 5, tanto para dados obtidos pela monitorização intensiva de utilizadores como para dados recolhidos a partir da publicação intermitente nas redes sociais, demonstrou-se que esta estrutura tem os conceitos necessários para a representação da mobilidade humana em ambos os cenários, ficando ainda demonstrado que a informação estruturada de acordo com esta representação permite a criação de diferentes vistas da mobilidade a partir de diferentes ferramentas de visualização.

Além de não existir um consenso comum para a definição da estrutura base de informação de mobilidade, como Yu (Yu e Shaw, 2004) refere, as ferramentas mais usadas no planeamento do espaço urbano (GIS) não têm capacidade de resposta adequada à componente temporal existente nos dados obtidos pela observação da mobilidade dos artefactos. Nesse sentido, um outro contributo deste trabalho passou pela definição dos conceitos com suporte para dados espaçotemporais. Assim, todos os fenómenos associados à mobilidade humana, com correspondência nos vários conceitos definidos pela estrutura, têm na sua representação, informação tanto espacial como temporal.

Ao nível espacial é de salientar ainda um contributo relativamente ao estado da arte, com a adição e distinção entre o posicionamento simbólico e o posicionamento geométrico dos artefactos. Com a inclusão de múltiplos sensores na observação dos artefactos houve necessidade de contemplar na estrutura de conceitos duas formas distintas de representação da posição, posição esta diretamente relacionada com a tecnologia usada para a observação.

É de salientar ainda que o modelo probabilístico implementado na validação da estrutura de conceitos permitiu concluir que ao adicionarmos mais sensores, e como tal mais dados relativos ao mesmo artefacto, os resultados melhoram e como tal a nossa estrutura de conceitos é nesse ponto de vista também um contributo relevante para o estado da arte.

Outro resultado relevante deste trabalho está descrito no capítulo 3. Este contributo não é tanto ao nível da investigação, mas mais uma visão futura do que poderia ser o Ecossistema Urbano, no entanto muito relevante para a problemática associada à mobilidade humana. Estudar padrões de mobilidade, formas de visualização e procurar explicações para a origem dos problemas associados à mobilidade será sempre um esforço inglório se não existirem equipas multidisciplinares compostas por elementos que analisem os dados e produzam resultados, bem como elementos responsáveis pelo planeamento e decisão sobre as alterações no espaço urbano. Com a constante mutação do espaço urbano e seus intervenientes, este ecossistema composto então por estas equipas deveria não só trabalhar em cooperação como em permanência, tentando avaliar e testar novas hipóteses bem como estudar os efeitos produzidos pelas alterações no espaço urbano.

Com a realização das duas experimentações apresentadas no capítulo 5 verificamos um fenómeno digno de relevância. Ao compararmos os resultados de análise de mobilidade, primeiro obtidos a partir da monitorização contínua de indivíduos com recurso a um smartphone, com os resultados obtidos posteriormente a partir de dados recolhidos de redes sociais, verifica-se que as conclusões extraídas a partir de ambas as análises de mobilidade têm importância e peso diferente para quem pretender com elas planear alterações no espaço urbano. Enquanto no primeiro caso (seção 5.2) é possível “seguir” os indivíduos de forma mais frequente e como tal mais próximo do real movimento humano, no segundo caso (seção 5.3) devido à publicação esparsa nas redes sociais é necessário ter alguns cuidados e reservas relativamente às conclusões que são inferidas a partir da análise de mobilidade realizada apenas com recurso a esta forma de observação do movimento humano.

1.3. Organização do documento

Este documento encontra-se estruturado por capítulos e respetivas secções e subsecções, onde neste primeiro capítulo o foco é a descrição do enquadramento do problema a ser alvo de trabalho ao longo do documento e os resultados que foram atingidos.

No capítulo 2 são apresentados alguns exemplos existentes na literatura, entre os quais as áreas de aplicação do estudo da mobilidade humana, como se caracteriza o movimento humano, diferentes formas de visualização da mobilidade quer seja individual ou global (coletiva), diferentes tipos de dados recolhidos a partir da observação dos artefactos, as técnicas mais comuns de análise de mobilidade e por último os desafios que ainda existem relativamente ao estudo da mobilidade humana em si e das técnicas e ferramentas necessárias para esse estudo.

O capítulo 3 descreve uma visão do que seria idealmente o Ecossistema Urbano para apoio e decisão sobre a mobilidade humana, um ecossistema baseado em equipas multidisciplinares de observação e ação sobre o espaço urbano. Descreve ainda uma proposta para a adequada realização da análise de mobilidade retirando partido da informação que consta nos *mapas individuais de mobilidade* complementada pela informação extraída da agregação destes mapas individuais em *mapas globais* ou dos mapas globais produzidos a partir de dados agregados.

O capítulo 4 apresenta a estrutura de conceitos desenvolvida para o mapeamento dos dados associados à mobilidade humana. Esta estrutura, composta por diferentes conceitos e processos de transformação, visa a homogeneização da informação de mobilidade independentemente da tipologia de dados usada na observação e nas técnicas ou formas de visualização da mobilidade.

O capítulo 5 contém duas implementações dos conceitos e processos de transformação da estrutura desenvolvida com tipologia de dados distinta permitindo a validação da estrutura apresentada no capítulo 4. As duas implementações permitem ainda verificar as diferenças nas conclusões que são possíveis de extrair a partir de monitorização contínua de indivíduos ou de dados recolhidos a partir de redes sociais.

Finalmente, no sexto e último capítulo são apresentadas as conclusões obtidas com o trabalho desenvolvido e são deixadas algumas propostas para trabalho futuro de investigação.

1.4. Contributos originais

Como resultado do trabalho desenvolvido e apresentado nos próximos capítulos desta tese, em Junho de 2012 foi defendido publicamente na conferência ICCSA em Salvador da Bahia, Brasil, um artigo intitulado *Dealing with multiple source spatio-temporal data in urban dynamics analysis* (Peixoto e Moreira, 2012) onde pela primeira vez se apresentou a estrutura de conceitos assim como o algoritmo de *place learning* baseado num modelo probabilístico para a integração de observações de diferentes sensores: GPS, Wi-Fi e GSM.

Posteriormente em 2013, e como resultado da implementação apresentada na seção 5.2, foi publicado um segundo artigo na revista *International Journal of Agricultural and Environmental Information Systems* intitulado *Human movement analysis using heterogeneous data sources* (Peixoto e Moreira, 2013) onde o foco principal é a validação da implementação dos vários processos de transformação da estrutura de conceitos e possibilidades de representação visual com base nos resultados obtidos no mapeamento dos dados nos diferentes conceitos.

Em Outubro de 2016, foi submetido à revista *Journal of Ambient Intelligence and Smart Environments*, o artigo *A Survey on Urban Mobility Analysis* cujo objetivo visa a apresentação de uma revisão bibliográfica sobre as diferentes perspetivas relacionadas com a observação e análise da mobilidade humana.

Igualmente em Outubro de 2016, foi ainda submetido à conferência *AGILE 2017*, o artigo *Limitations on Human Mobility Analysis from Social Networks Data* cujo tema aborda a utilização de dados recolhidos a partir de redes sociais e o seu potencial de utilização na análise de mobilidade humana, procurando-se extrair indicadores sobre que conclusões se podem ou não inferir a partir das diferentes análises que se obtêm com base nesta fonte de dados.

Capítulo 2. Perspetivas sobre a mobilidade humana

A mobilidade humana despertou e desperta o interesse da comunidade científica em diferentes perspetivas. Se por um lado alguns autores aceitaram o desafio de tentar reproduzir, com modelos probabilísticos ou outras representações, os padrões comportamentais dos movimentos humanos, outros autores procuraram formas tecnológicas de recolher informação de mobilidade de artefactos (quer sejam indivíduos; equipamentos associados aos indivíduos, como os smartphones; ou mesmo outros objetos que possam ser usados como *proxies*¹ da mobilidade dos indivíduos).

Por existirem já muitos trabalhos e diferentes perspetivas, neste capítulo serão apresentados alguns exemplos desses trabalhos como forma de ilustrar de forma alargada as diferentes vertentes associadas à mobilidade humana.

Este capítulo encontra-se assim organizado em 6 seções: áreas aplicacionais da análise da mobilidade; a natureza do movimento humano; diferentes formas de visualização do movimento humano; diferentes tipos de sensores e formas de observação; algumas das técnicas de análise e, por último, os desafios ainda sem resposta.

¹ O conceito de *proxy* neste trabalho refere-se a objetos ou aplicações que por acompanharem as pessoas no seu dia-a-dia refletem através da sua observação uma realidade ao nível da mobilidade muito semelhante à mobilidade desses indivíduos.

2.1. Áreas aplicacionais

Em 1995, Louis Alfeld (Alfeld, 1995) considerava que num futuro próximo (que corresponde à data atual) a observação das dinâmicas urbanas deveria auxiliar os decisores ou planeadores das cidades e outros espaços urbanos. A razão principal prendia-se com a visão que tinha para o ano 2020, onde a população mundial, poluição e recursos naturais iriam, segundo o autor, atingir valores problemáticos sendo por isso importante o estudo das dinâmicas urbanas de forma a serem identificados os problemas e através da sua identificação procurarem-se soluções para os resolver ou, pelo menos, minimizar os seus efeitos.

Nos últimos anos vários trabalhos têm procurado atingir esse objetivo - analisar as dinâmicas das cidades – de forma a auxiliar os decisores na sua tomada de decisão e/ou revelar situações anómalas nas cidades que de outra forma poderiam não ser visíveis de forma simples.

Estas análises estão direcionadas para diferentes áreas aplicacionais. A literatura apresenta já um conjunto alargado e nesta secção temos uma amostra de algumas dessas áreas aplicacionais como ilustração da aplicabilidade da análise de dinâmicas humanas.

A deteção de todo o tipo de situações de emergência (Mathioudakis e Koudas, 2010) ou, apenas situações de catástrofes (por exemplo: terremotos e furacões) (Sakaki, Okazaki e Matsuo, 2010), ou ainda a gestão posterior de situações de crise em resultado de emergências (MacEachren e Robinson, 2011) são algumas das várias aplicações da observação e interpretação de fenómenos existentes nos espaços urbanos.

Mas a segurança de uma cidade não são apenas situações anómalas de emergência ou catástrofe. É também importante para o decisor saber quais as zonas da cidade onde existe, por exemplo, uma maior probabilidade de ocorrência de alteração da ordem pública. Por este motivo, uma área de interesse para a análise de dinâmicas da cidade é a compreensão e mapeamento de zonas com comportamentos de risco (Young, Rivers e Lewis, 2014) ou zonas da cidade onde existe maior ou menor atividade em diferentes momentos do dia (propiciando esse tipo de situações), recorrendo por exemplo à análise de ruído existente na própria cidade (D'Hondt, Stevens e Jacobs, 2013).

A análise de atividade em diferentes zonas da cidade pode ser focada em determinadas características como, por exemplo, a deteção de eventos culturais em zonas geograficamente reduzidas (praças, ruas, ...) (Xia *et al.*, 2014). A extração de informação sobre atividades/eventos pode inclusivamente ser usada pelo decisor para identificar áreas geográficas com comportamentos

distintos: zonas residências; zonas comerciais/industriais; ou ainda zonas onde a atividade mais intensa ocorre durante a noite (Frias-Martinez e Frias-Martinez, 2014).

Até ao momento vimos como exemplo duas grandes áreas aplicacionais da observação da dinâmica das cidades, no entanto não podemos esquecer a área que mais tem cativado os investigadores nos últimos tempos: o planeamento urbano.

Esta grande área aplicacional é de verdadeira importância, pois como refere Díaz (Díaz-Palacios-Sisternes, Ayuga e García, 2014) a dinâmica das cidades tem que ser analisada na perspetiva de uma relação entre os fenómenos urbanos e as infraestruturas. Por esse motivo o planeamento de atuais e novas infraestruturas tem de ser sempre analisado de acordo com as atuais dinâmicas existentes nas cidades.

O planeamento na criação ou alteração de infraestruturas pode ser feito num âmbito mais alargado quer ao nível do tempo, quer ao nível do espaço, como Díaz (Díaz-Palacios-Sisternes, Ayuga e García, 2014) fez para compreenderem a evolução das áreas urbanas através do aparecimento ou desaparecimento de zonas rurais e habitacionais. Este tipo de análise pretende descrever como foi o crescimento urbano e/ou que modificações se verificaram na forma de uso do terreno (Nassar, Alan Blackburn e Duncan Whyatt, 2014), possibilitando desta forma uma melhor planificação da potencial evolução futura do espaço urbano.

A evolução do crescimento urbano, como área aplicacional, permite ao decisor ter uma base de conhecimento interessante sobre a forma como os espaços urbanos cresceram/diminuíram e quais as transformações no uso/ocupação desse mesmo espaço. No entanto, quando se pretende instalar novas infraestruturas numa cidade é muito importante conhecer as necessidades dos cidadãos que utilizam diariamente essas infraestruturas. Como Andrienko (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013) refere, o planeamento dos espaços urbanos deve ter em consideração os movimentos e comportamentos existentes nestes mesmos espaços, sempre com o objetivo de adequar a melhor resposta ao nível do planeamento às necessidades dos habitantes desses espaços.

No entanto, muitas vezes essa informação sobre movimentos e comportamentos não chega sequer aos decisores, porque na maioria dos casos a mesma simplesmente não existe. Por esse motivo

tem-se verificado o aparecimento de aplicações colaborativas, como é o caso da *WeCycle*², cujo objetivo é conhecer as necessidades dos ciclistas de forma a ajudar no planeamento de novas infraestruturas (neste caso em particular no planeamento de ciclovias para o município de Londres).

Mas não é só no planeamento de novas infraestruturas que a observação da mobilidade e o seu mapeamento são importantes para o decisor do espaço urbano. Uma outra área aplicacional muito importante no planeamento está associada aos problemáticos congestionamentos de tráfego, que são, infelizmente, atualmente fenómenos “naturais” das médias e grandes cidades de todo o mundo, com elevados custos económicos e que afetam negativamente a qualidade de vida das populações que diariamente têm que lidar com este tipo de fenómeno.

Assim a identificação e compreensão de fenómenos de congestionamento de tráfego tem sido a área aplicacional a que a comunidade científica tem dado mais destaque, seja através do recurso às redes sociais (Ribeiro *et al.*, 2014) ou à análise de dados de GPS (Gennady Andrienko *et al.*, 2013).

Este tipo de observação de mobilidade permite a identificação de diferentes tipos de situações: determinar zonas de congestionamento de tráfego (Andrienko *et al.*, 2011), ou períodos do dia em que o trânsito tem comportamentos anómalos (Andrienko *et al.*, 2010).

Enquanto alguns dos trabalhos focam a mobilidade ao nível da cidade, Liu (Liu *et al.*, 2014) tenta extrair padrões de mobilidade coletiva entre grandes cidades, onde o objetivo não é tanto a consequência (congestionamentos de tráfego) mas sim a origem, isto é, compreender entre que cidades existe uma forte mobilidade e quais as cidades que desta forma têm uma maior interação.

Em conclusão, existem neste momento um conjunto relevante de áreas aplicacionais da observação da dinâmica das cidades, tendo sido apresentadas nesta secção apenas alguns desses exemplos. A Figura 2.1 resume as várias áreas aplicacionais anteriormente referidas, existindo, no entanto, outros exemplos.

² <http://www.fastcoexist.com/3034313/a-new-app-tracks-where-cyclists-actually-ride-to-help-plan-better-paths> (visitado em Agosto de 2016)

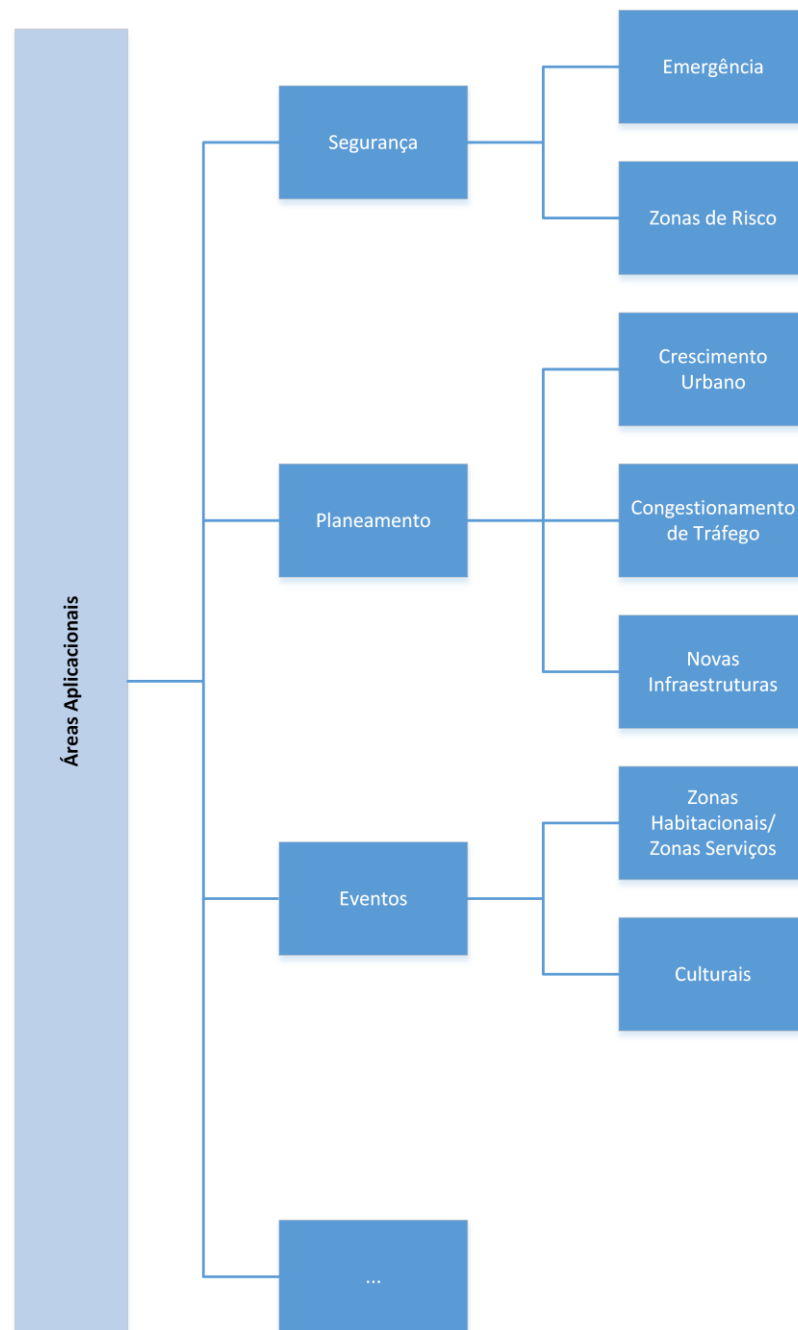


Figura 2.1 - Exemplos de algumas Áreas Aplicacionais

2.2. Natureza do movimento humano

Com base nas vivências e crenças do nosso dia-a-dia verificamos que o homem é um ser de hábitos. Tem por hábito realizar os mesmos trajetos entre, por exemplo, casa-trabalho e trabalho-casa; visitar com alguma frequência locais com os quais mais se identifica ou pelos quais tem maiores preferências; realizar um conjunto de rotinas nos dias de semana e outras apenas aos fins de semana;

entre outros exemplos de comportamentos que com maior ou menor frequência se repetem ao longo do tempo.

Por este motivo, o conhecimento baseado no senso comum leva-nos a considerar que as características humanas e o carácter repetitivo com que vivemos o nosso dia-a-dia, mantendo uma coerência ao nível do nosso comportamento espaçotemporal, pode e é uma forma muito interessante de, com base no passado, reproduzirmos ou prevermos de alguma forma o futuro. Observando as movimentações que realizámos num passado recente, podemos assim extrapolar quais poderão ser as movimentações num futuro próximo desde que mantidas as mesmas condições.

Mas o conhecimento baseado no senso comum nem sempre coincide com o conhecimento científico e apenas a nossa experiência diária não é suficiente para podermos generalizar os padrões da mobilidade humana.

No trabalho realizado por Brockmann (Brockmann, Hufnagel e Geisel, 2006) as conclusões foram precisamente contrárias a esta hipótese. Com base no estudo dos movimentos de notas bancárias ao longo dos Estados Unidos, as conclusões a que chegaram relativamente ao movimento humano indicavam que este apresentava um comportamento aleatório (*random walk model*) não sendo por isso conclusões animadoras quanto à possibilidade de podermos explicar e/ou replicar, com um grau de certeza elevado, as movimentações humanas futuras com base na análise do passado.

Posteriormente González (González, Hidalgo e Barabási, 2008) tentou com outra abordagem ao nível dos dados (dados sobre a utilização de rede móveis celulares), avaliar se as trajetórias humanas seriam então aleatórias ou se, pelo contrário, apresentavam alguma regularidade ou padrão. Como resultado deste trabalho, publicado na revista Nature, demonstraram que existe uma forte tendência para o ser humano retornar aos locais anteriormente visitados de forma recorrente e com alguma periodicidade temporal; verificaram ainda que existe um conjunto reduzido de locais onde passamos a maior parte do tempo sendo que dois deles são normalmente a nossa casa e o local de trabalho. Verificaram também que nos dois locais mais visitados os utilizadores passavam cerca de 40% do seu tempo e que os locais mais frequentemente visitados seriam cerca de 5, pelo que os restantes locais (que poderiam chegar aos 50 locais distintos) eram visitados com uma frequência que ia regularmente diminuindo (Figura 2.2).

Estas conclusões permitem dar uma indicação de que o movimento humano pode ser reproduzido e que o mesmo segue uma distribuição probabilística.

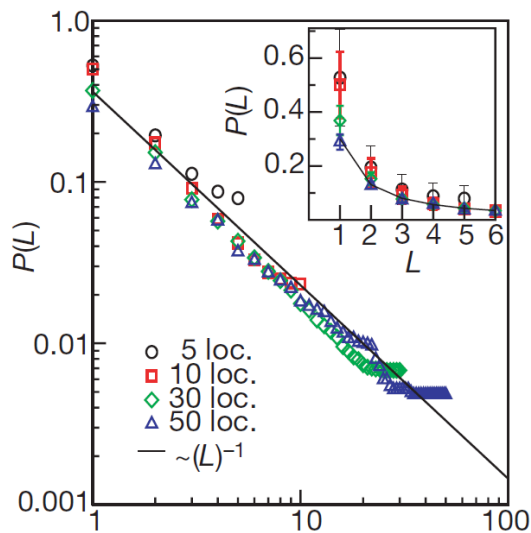


Figura 2.2 - Frequência das visitas aos locais (González, Hidalgo e Barabási, 2008)

A Figura 2.2 descreve o padrão de visitas a locais distintos por quatro diferentes utilizadores, representados por diferentes símbolos. O eixo horizontal representa o número de diferentes locais visitados L de acordo com a frequência de visita, isto é, do 1 ao 5 estão os cinco locais mais visitados ordenados de acordo com a frequência de visita. No eixo vertical encontra-se o cálculo da probabilidade do utilizador $P(L)$ se encontrar num desses locais. Assim a probabilidade $P(L)$ de um utilizador se encontrar num determinado local pode ser descrita de forma aproximada por $1/L$ (reta descendente apresentada na figura), o que indica que o aumento de locais visitados influencia naturalmente a probabilidade de acerto relativamente à localização do utilizador. Complementarmente, o gráfico de dimensões mais reduzidas apresentado na mesma figura descreve a forma como os indivíduos permanecem no tempo nesses locais. De acordo com este estudo, os indivíduos passam a maior parte do tempo nos locais com maior frequência de visita.

Posteriormente, em 2010, Song (Song *et al.*, 2010) pretendeu demonstrar que não só o comportamento humano ao nível da mobilidade é um fenómeno previsível, como é possível quantificar essa previsibilidade, chegando à conclusão que se consegue prever, com uma certeza acima dos 80%, quais os locais visitados pelos utilizadores. Este estudo releva ainda, em adição aos resultados obtidos por González (González, Hidalgo e Barabási, 2008), que dois dos locais com maior número de visitas (segundo os autores muito possivelmente correspondentes à casa e ao local de trabalho) têm uma previsibilidade de serem visitados a rondar os 60% e que a previsibilidade aumenta até muito próximo dos 80% para os 5 locais mais visitados (Figura 2.3).

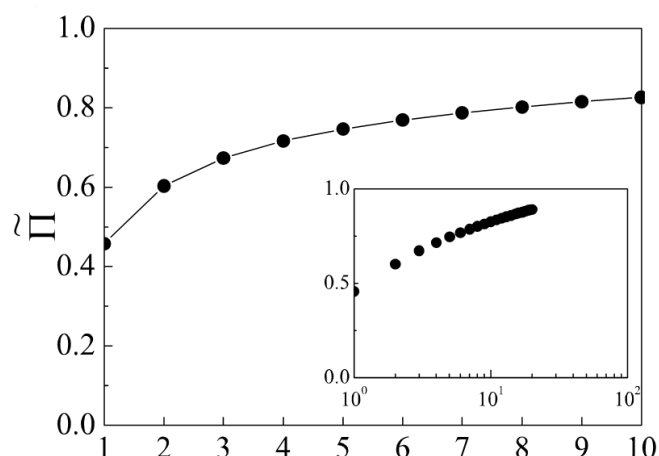


Figura 2.3 – Relação entre os 10 locais mais visitados e a sua previsibilidade (Song et al. 2010)

Nos resultados obtidos por González (González, Hidalgo e Barabási, 2008) e atestados por Song (Song *et al.*, 2010) verifica-se a existência de um valor indicativo - 5 locais distintos - como os locais mais visitados devido à frequência da visita e como os locais onde o tempo de presença é também predominante.

Comparativamente com os resultados obtidos por Brockmann (Brockmann, Hufnagel e Geisel, 2006) verificam-se conclusões distintas se não mesmo contraditórias quanto à possibilidade de ser ou não previsível o movimento humano. Uma possível explicação para as diferenças nas conclusões prende-se com a escala espacial e temporal da análise realizada. Enquanto no trabalho de Brockmann (Brockmann, Hufnagel e Geisel, 2006), quando se estudam as trajetórias de curta distância, o intervalo de tempo entre a origem e destino da trajetória é de um a quatro dias, no caso do trabalho de González (González, Hidalgo e Barabási, 2008) estes intervalos de tempo são reduzidos (um dos *datasets* usado tinha intervalos de tempo de amostragem igual ou inferior a duas horas). Desta forma, no trabalho de Brockmann (Brockmann, Hufnagel e Geisel, 2006) é muito provável que os dados e consequente análise de mobilidade não reflitam assim todas as trajetórias realizadas durante o nosso dia-a-dia uma vez que o intervalo mínimo era de um dia. A escolha do tipo de dados usados na análise da mobilidade humana condiciona o período de amostragem com que os dados são recolhidos e como tal estamos perante cenários de mobilidade distintos (ao nível da escala espacial e temporal) o que poderá ser um dos motivos para a diferença nas conclusões destes dois trabalhos.

O trabalho realizado por Lind (Lind e Moreira, 2015) é um exemplo que corrobora esta hipótese. Neste trabalho, onde a área geográfica do estudo foca um espaço territorial reduzido, com trajetórias e respetivas distâncias numa escala muito menor quando comparadas com os trabalhos

anteriores e onde a duração das trajetórias é por isso também reduzida, os resultados obtidos vão no sentido de que a distância das trajetórias é um fator importante na caracterização da mobilidade. Uma das conclusões deste trabalho refere que o tipo de escala da análise (se alargada ou mais reduzida) afeta a caracterização da tipologia de mobilidade uma vez que aumentando a dimensão geográfica da análise dá origem ao aparecimento de trajetórias mais longas e com maior frequência, contrariamente ao que acontece em área geográficas mais reduzidas onde essas trajetórias são em número reduzido ou inexistentes. Relativamente aos padrões de mobilidade, neste trabalho é demonstrado que as movimentações em espaços pequenos tendem a seguir uma distribuição Browniana, onde as movimentações são aleatórias, em semelhança ao descrito no trabalho de Brockmann (Brockmann, Hufnagel e Geisel, 2006).

Na mesma linha do trabalho anterior, Krumm (Krumm, Caruana e Counts, 2013) observa que uma característica importante na escolha dos locais que visitamos é a distância a que nos encontramos desse local. De um conjunto de 61 características usadas para descrever uma zona geográfica, entre elas fatores demográficos como: a idade, a raça, a média de idades por habitação, entre outras, a característica com maior peso na decisão sobre o próximo local a visitar é a distância entre esse local e a nossa habitação. O algoritmo de classificação implementado nesse trabalho obteve uma taxa de acerto de 91% quando se considerou apenas para o treino e teste do algoritmo a característica distância. Se adicionadas as 61 características, a percentagem de acerto subia ligeiramente para os 95%.

Ainda no trabalho de Brockmann (Brockmann, Hufnagel e Geisel, 2006) existe uma outra conclusão igualmente interessante para descrever a mobilidade humana: a frequência com que se realizam as trajetórias diminui com o aumento da distância (Figura 2.4). À medida que o deslocamento relativamente a um ponto inicial aumenta, a frequência com que se realiza esse deslocamento diminui. Com base no número de movimentações que foram realizadas pelas notas bancárias, os autores constataram que até ao máximo de 10 kms de raio, a probabilidade de se observar uma movimentação é bastante alta e que a partir desta distância/deslocamento, a probabilidade de se observar uma movimentação tende a diminuir à medida que a distância entre locais aumenta.

Esta conclusão é particularmente útil para a análise de mobilidade ao nível da cidade, uma vez que reforça a ideia presente no senso comum de que os locais mais visitados estão geograficamente próximos e como tal as distâncias percorridas entre esses locais são relativamente curtas. Como

podemos verificar pela Figura 2.4, a partir dos 10 km de raio, o número de trajetórias realizadas inicia uma fase descendente de probabilidade de ocorrência.

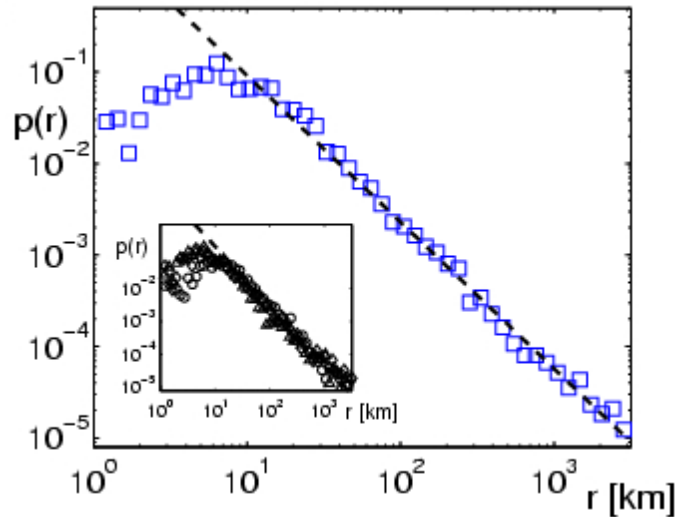


Figura 2.4 – Distribuição das distâncias (Brockmann, Hufnagel e Geisel, 2006)

Os diferentes trabalhos descritos anteriormente, embora possam ter conclusões contrárias entre eles, em virtude não do trabalho de investigação, mas sim dos dados à disposição para a realização dos estudos, permitem resumir algumas características interessantes relativas à natureza do movimento humano:

- Existem cerca de 5 locais de relevância para cada cidadão;
- Os locais mais relevantes tendem a ficar circunscritos a um raio de 10 km de distância;
- A caracterização da mobilidade humana é influenciada pelas distâncias entre a origem e o destino;
- O comprimento das deslocações entre os diferentes locais visitados por um indivíduo segue uma distribuição que pode ser aproximada por uma lei de potências truncada (*truncated power law*);
- Independentemente do modelo probabilístico, existem indícios da previsibilidade na natureza do movimento humano.

2.3. Visualização do movimento humano

Porque “*uma imagem vale mais do que mil palavras*”, a análise da mobilidade humana recorre frequentemente a representações, estáticas ou dinâmicas, dos movimentos humanos, quer estes sejam individuais ou coletivos.

Uma simples representação usando **séries temporais**, como Reades (Reades *et al.*, 2007) fez para descrever a **ocupação do espaço** ao longo do dia e em diferentes dias para 6 locais distintos, é um bom ponto de partida para extrair conclusões sobre quais os locais mais povoados por indivíduos e em que momentos temporais os mesmos ocorrem (Figura 2.5).

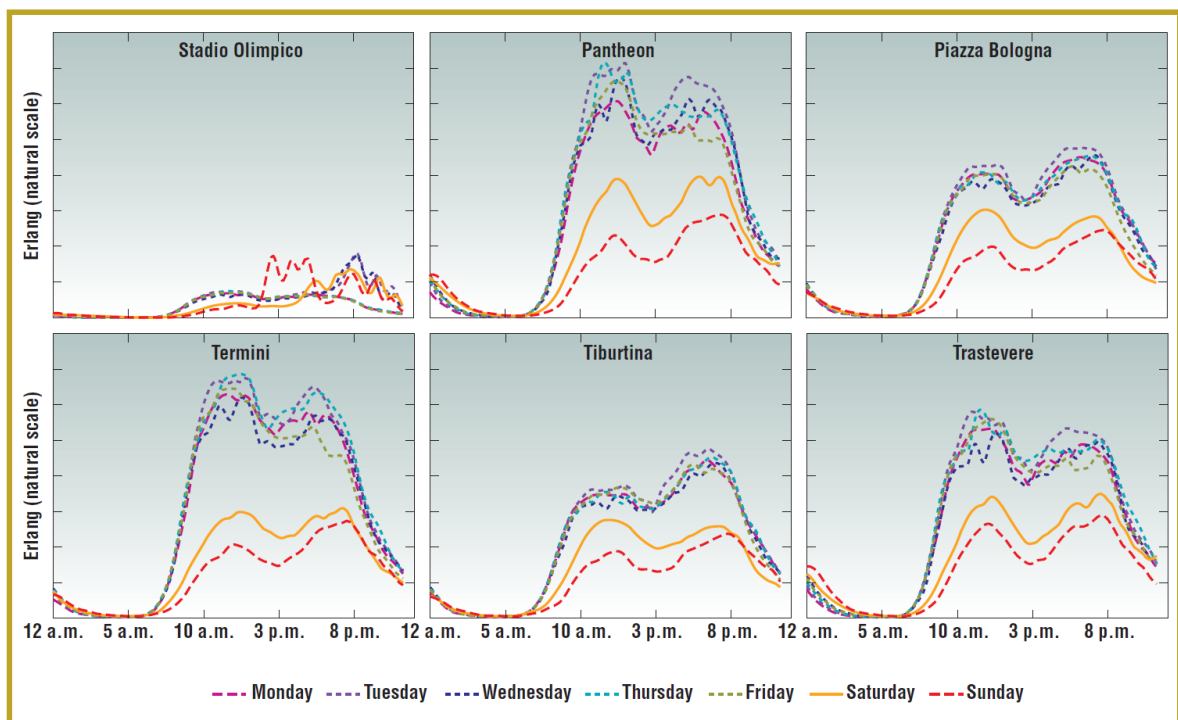


Figura 2.5 – Visualização da ocupação do espaço ao longo do tempo (Reades et al., 2007)

No entanto, embora a representação anterior já transmita indicadores importantes sobre a evolução da ocupação do espaço urbano ao longo do tempo, falta-lhe uma componente geográfica de forma a contextualizar a informação. Por este motivo, de forma complementar ou em substituição destas séries temporais, são usadas frequentemente visualizações com base geográfica que ilustram a distribuição dos indivíduos ao longo do espaço urbano (Figura 2.6).

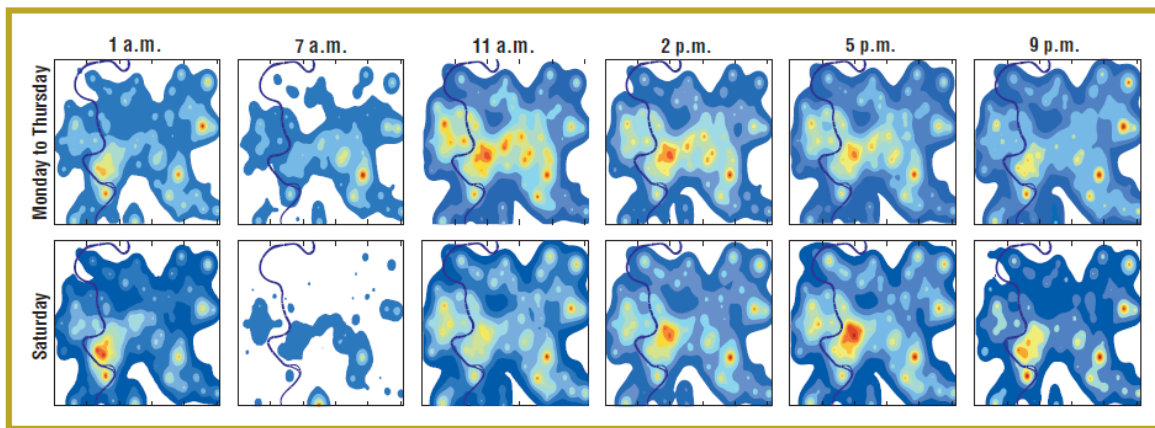


Figura 2.6 – Visualização da cidade de Roma em diferentes horas do dia (Reades et al., 2007)

Aparecem assim visualizações do movimento humano com recurso a imagens estáticas que descrevem o estado momentâneo da ocupação do espaço por parte dos indivíduos. De salientar que a representação da ocupação de espaço apenas indica que um conjunto de indivíduos foi observado em determinado local durante um certo intervalo de tempo, não ilustra a movimentação dos indivíduos mesmo que tenhamos uma sequência de imagens como acontece na Figura 2.6. Existe uma diferença significativa em analisar padrões de ocupação de espaço e padrões de movimentações. Enquanto no primeiro caso temos a identificação de locais de maior ou menor densidade espacial de pessoas, no segundo caso o foco é a identificação de trajetórias mais ou menos frequentes com base no início e fim do movimento.

Estas representações, conhecidas por **snapshots**, são uma “fotografia” do estado da cidade apenas no momento (um período muito curto de tempo) sendo por este motivo comum serem extraídos vários *snapshots* em diferentes momentos temporais permitindo desta forma ter uma sequência de imagens simulando desta forma a dinâmica existente, como acontece com a figura anterior.

A utilização de *snapshots*, mesmo sendo das primeiras abordagens a surgirem como metodologia de visualização de dinâmicas humanas, é ainda hoje uma das opções com grande utilização na ilustração da ocupação do espaço urbano.

De forma a tornar mais legível a informação representada pelos *snapshots* é comum a utilização de uma representação gráfica denominada de **heatmap** (Figura 2.6), onde com o recurso à codificação das cores é feita uma classificação dos dados apresentados. Na figura anterior, por exemplo, o azul e verde representam “zonas frias” com menores ocorrências, enquanto o laranja e vermelho representam “zonas quentes” onde as ocorrências são elevadas (neste caso, as cores representam a densidade espacial de pessoas).

Com recurso a esta técnica de representação gráfica torna-se mais intuitiva a compreensão de, por exemplo, zonas de maior concentração de pessoas. A título de exemplo, se compararmos a Figura 2.6 e a Figura 2.7 verificamos que ambas representam a distribuição de indivíduos ao longo do espaço, embora apenas na primeira figura seja evidente a supremacia com que determinado espaço está a ser ocupado relativamente aos outros locais em análise.

No entanto, não é só pelo facto de termos ou não uma classificação dos dados que estas duas figuras se distinguem. São também distintas relativamente ao período de análise que representam.

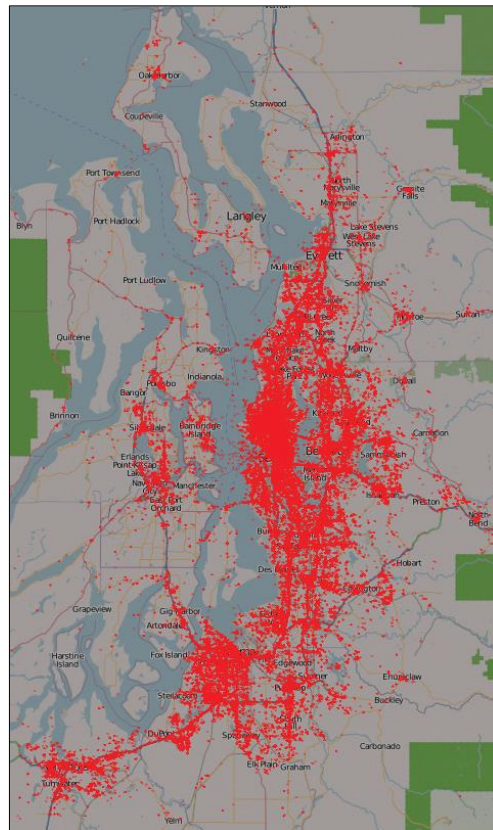


Figura 2.7 - Visualização de zonas com maior concentração de pessoas (G. Andrienko, Andrienko, Bosch, et al., 2013)

Enquanto a Figura 2.6 representa informação momentânea, resultante da análise de períodos temporais muito curtos, já a Figura 2.7 representa um agregado de informação, neste caso em particular relativo a um dia de *tweets* publicados. Não seria por isso correto apelidar esta forma de visualização também de *snapshot* fazendo por isso mais sentido apelidar a mesma de ***footprint***.

A opção por esta terminologia prende-se pelo facto de que cada indivíduo, através do que publicou na rede social *Twitter*, ir ao longo do tempo deixando no espaço urbano a sua pegada eletrónica composta por um rasto de publicações realizadas na rede social em questão.

Temos assim na figura anterior um *footprint* global, proveniente de vários indivíduos, ao longo de um dia circunscrito a uma área geográfica.

A distinção entre a terminologia *snapshot* e *footprint* é então feita de acordo com o período temporal da análise em questão. Em períodos temporais reduzidos (até ao máximo de algumas horas) a visualização apresenta o momento, enquanto ao representarmos por exemplo um dia ou mais de movimentações, estamos perante um resumo/acumulado do que aconteceu e não o estado momentâneo do espaço urbano.

Por sua vez, a utilização de *heatmaps* como representação gráfica classificadora de dados está presente em várias técnicas de visualização, não sendo exclusiva nem dos *snapshots*, nem dos *footprints*, apresentados anteriormente.

Nos exemplos da Figura 2.6 e Figura 2.7 as visualizações ilustram a ocupação do espaço urbano num período de tempo e espaço. Mas a ocupação do espaço urbano não é só descrita pela “quantidade de pessoas por metro quadrado”. Jiang (Jiang, Ferreira e Gonzalez, 2012) recorre também a *heatmaps* para descrever a ocupação do espaço mas de acordo com o tempo que as pessoas passam em determinados locais. A Figura 2.8 representa a ocupação do espaço de acordo com os tempos de estadia nos locais, isto é, as zonas “quentes” do *heatmap* referem-se a locais com forte presença temporal dos indivíduos. Não significa que haja muitos indivíduos nesses locais, mas sim, que esses indivíduos passam nesse local muito do seu tempo diário quando comparado com as restantes estadias.



Figura 2.8 – Densidade temporal (Jiang, Ferreira e Gonzalez, 2012)

A título meramente ilustrativo, comparativo das representações anteriores, a nossa casa ou o nosso local de trabalho seriam com grande probabilidade dois locais com forte presença temporal (uma zona “quente” semelhante às da Figura 2.8 no nosso mapa individual de mobilidade) no entanto não significa que o sejam relativamente à perspetiva representada na Figura 2.6.

Da mesma forma que a ocupação do espaço poderá ser descrita, ou pela quantidade de indivíduos por área geográfica, ou pela quantidade temporal da estadia numa área geográfica, as visualizações apresentadas anteriormente podem ser utilizadas para mais do que a representação da ocupação do espaço.

Andrienko (Andrienko *et al.*, 2010) utilizou uma representação ligeiramente diferente da combinação de *snapshots* com *heatmaps* que permite a visualização da velocidade média ao longo da cidade de Milão (que permitirá inferir, por exemplo, potenciais zonas de congestionamento de tráfego) com base numa representação geográfica incluindo também a componente temporal (Figura 2.9).

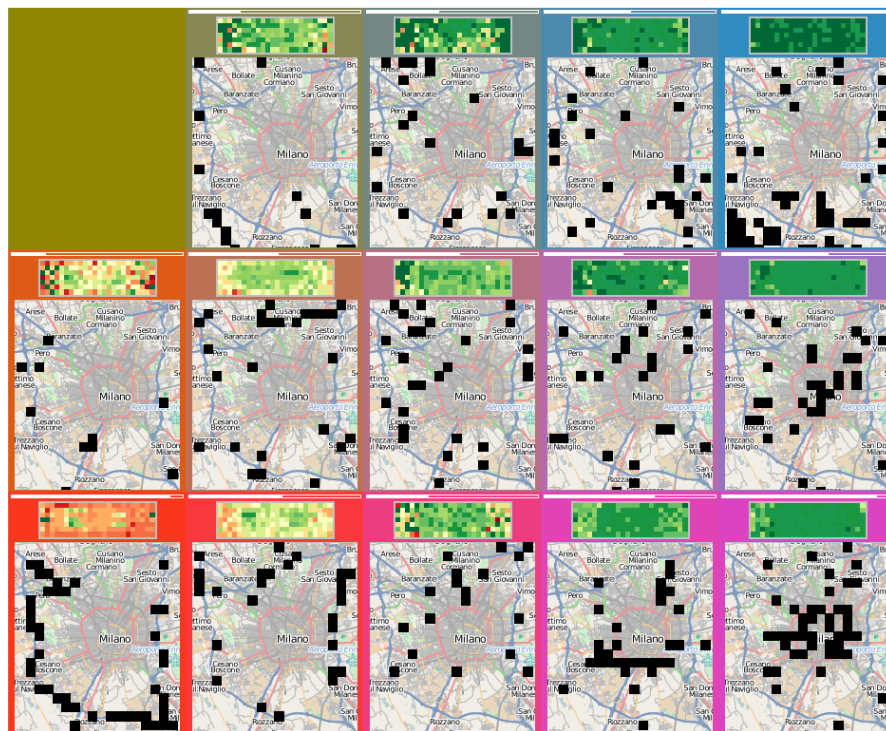


Figura 2.9 – Visualização de velocidades médias ao longo do tempo e no espaço (Andrienko et al., 2010)

Na Figura 2.9 estamos perante uma representação distinta da representação da ocupação do espaço urbano. A representação da ocupação do espaço recorre ou descreve principalmente o estado imóvel dos artefactos, isto é, representa os momentos em que os indivíduos se encontram imóveis num local (não se considera para o efeito os movimentos realizados dentro dos edifícios). No entanto, descrever onde o indivíduo está e por quanto tempo esse mesmo indivíduo está num local não é em si mobilidade. A mobilidade dos indivíduos acontece porque os indivíduos deslocam-se de um ponto para outro no espaço, através de um determinado percurso, com um determinado perfil de velocidade e ao longo de um certo tempo.

Embora na Figura 2.9 não sejam claros esses movimentos ou as trajetórias realizadas pelos indivíduos, estão nela representadas as consequências desses mesmos movimentos ou da mobilidade humana. A representação que Andrienko faz das velocidades médias dos veículos que circulam ao longo da cidade são resultado das movimentações dos indivíduos e consequentes implicações ao nível do tráfego.

A Figura 2.9 introduz uma necessidade, a de encontrar uma forma de representação da mobilidade humana que vá para além da ocupação do espaço. Uma das formas encontradas para descrever os movimentos entre os locais visitados pelos indivíduos passa pela representação de **trajetórias**.

Andrienko (Andrienko *et al.*, 2011) recorre à informação de GPS para ilustrar por um lado as trajetórias mais usadas pelos condutores numa cidade (Figura 2.10a) e por outro, com base na mesma informação, para inferir situações em que a velocidade é reduzida sendo por isso um potencial indicador de congestionamento de tráfego (Figura 2.10b).

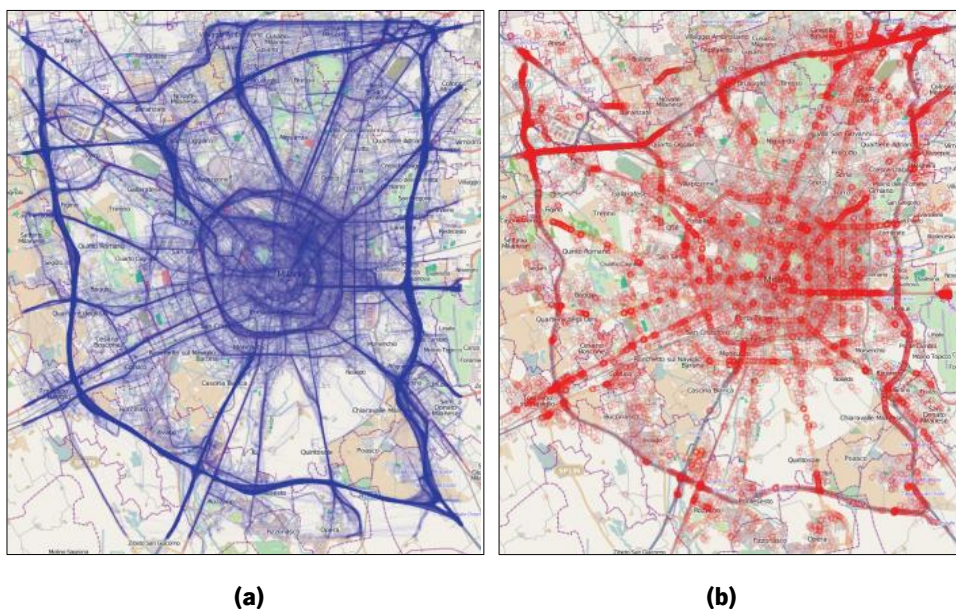


Figura 2.10- Trajetórias de veículos (a) e zonas de velocidade baixa (b) (Andrienko et al., 2011)

A Figura 2.10a apresenta um *footprint* coletivo das trajetórias realizadas pelos condutores numa cidade durante um dia. Esta representação não ilustra onde e quando as pessoas passam a maior parte do seu tempo ao longo do espaço urbano, mas sim, e sendo esta representação muito rigorosa, as vias mais utilizadas pelos condutores e que, como tal, potenciadoras de originar os congestionamentos de tráfego ilustrados na Figura 2.10b. De salientar que a Figura 2.10b não é em si uma representação propriamente dita de movimentos e consequentes trajetórias, mas sim uma

conclusão extraída desse movimento. No entanto com recurso a esta representação temos uma forma auxiliar de caracterizar o espaço, com a distinção de zonas de maior ou menor congestionamento. A Figura 2.10b enquanto representativa de um dia de análise de dados permite a obtenção de um *footprint* com o resumo das zonas mais críticas da cidade devido aos fortes congestionamentos. Se pelo contrário for realizada em intervalos de tempo de maior granularidade permite a produção de vários *snapshots* que permitem ilustrar como as zonas congestionadas variam ao longo do dia.

A Figura 2.10a é ainda um bom exemplo para a ilustração da noção de *footprint* e distinção relativamente aos *snapshots* uma vez que estamos perante um somatório de vários momentos (momentos esses que correspondem às trajetórias realizadas). Este somatório é neste caso demonstrado na intensidade de cor das trajetórias. As trajetórias com maior ocorrência ficam com maior definição e maior intensidade de cor, enquanto as trajetórias realizadas em menor número aparecem mais esbatidas na figura.

No exemplo da Figura 2.10a verificamos que a representação das trajetórias realizadas num espaço urbano torna mais evidente a dinâmica humana existente na cidade quando comparado com a análise através da ocupação de espaço.

Embora a Figura 2.10 faça uma representação geograficamente rigorosa das trajetórias, no entanto, nem sempre esse rigor poderá ser assegurado, existindo por isso exemplos de representações de trajetórias onde apenas a origem e destino das movimentações são representados com rigor (Figura 2.11).

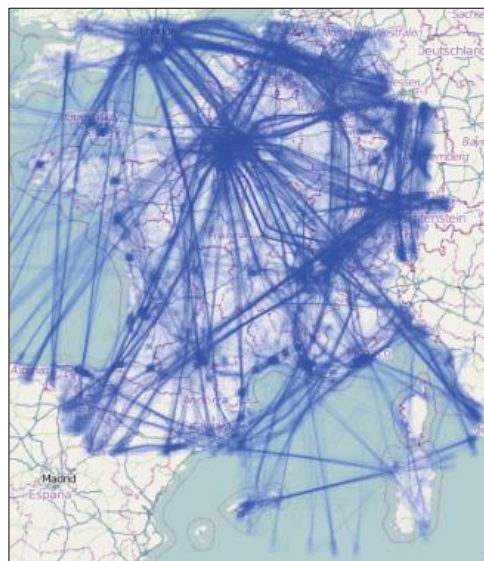


Figura 2.11 – Trajetórias de voos (Andrienko et al., 2011)

Na Figura 2.11 os autores (Andrienko *et al.*, 2011) ilustram trajetórias de voos realizados, neste caso em particular, em França e arredores. Nesta representação o rigor cinge-se apenas à localização dos aeroportos e aeródromos, sendo que a representação do movimento não é de modo algum fiel à realidade.

Esta figura, embora sendo também uma representação do movimento através de trajetória, representa a dinâmica de movimentos de uma forma diferente, com recurso à representação gráfica através de **grafos**.

Brockmann (Brockmann e Theis, 2008) apresenta uma visualização interessante da mobilidade humana recorrendo também à representação de trajetórias através de grafos (Figura 2.12). Com base na informação da circulação de notas³ ao longo dos Estados Unidos, e considerando que a mobilidade das notas está correlacionada com a própria mobilidade humana, fizeram o mapeamento dos movimentos das notas ao longo do espaço permitindo assim, em certa medida, inferir os movimentos humanos nesse mesmo espaço.

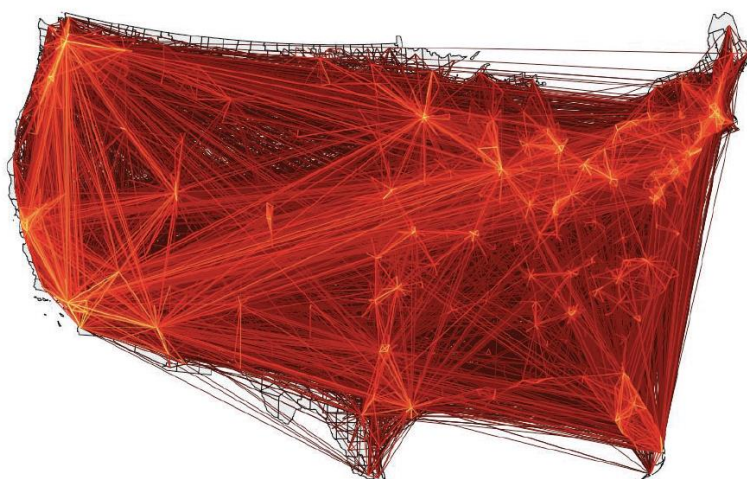


Figura 2.12 – Representação das trajetórias de notas (Brockmann e Theis, 2008)

O que torna interessante este trabalho não é a representação em si das trajetórias através de grafos, mas sim a informação que foi usada para a representação das mesmas. Em vez de serem usados dados de, por exemplo, dispositivos móveis, como *proxy* da mobilidade humana, neste caso foram utilizados os registos dos locais “visitados pelas notas”.

³ www.wheresgeorge.com (visitado em Agosto de 2016)

Song (Song *et al.*, 2010) faz uma representação de trajetórias através de grafos ligeiramente diferente da apresentada anteriormente (ver Figura 2.13).

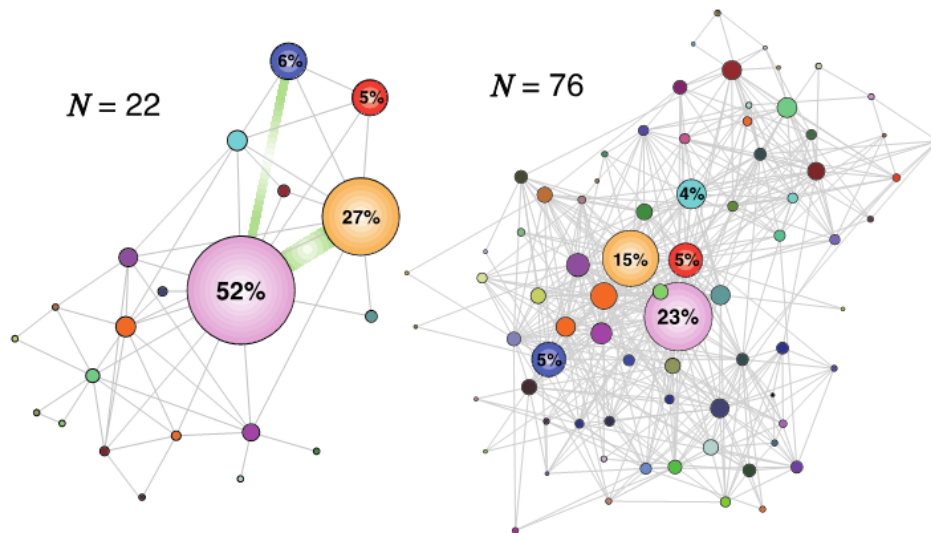


Figura 2.13 – Representação baseada em grafos (Song *et al.*, 2010)

Nesta figura são representados dois comportamentos de mobilidade humana para dois utilizadores respetivamente, onde os vértices dos grafos representam as células GSM onde foram realizadas chamadas pelos utilizadores (o diâmetro do vértice está associado ao valor numérico das ocorrências) e as arestas representam as transições entre as células GSM (onde a espessura das arestas é igualmente representativa do maior ou menor número de transições), que em certa medida refletem os movimentos realizados.

Com base nos exemplos anteriores de representação de trajetórias de mobilidade humana verificam-se duas formas de representação: geograficamente mais rigorosa (maior detalhe espacial), quando são utilizados dados de GPS para ilustrar as trajetórias; ou uma forma mais resumida de representação de trajetórias com base em grafos. Mais uma vez, como acontecia com a representação da ocupação de espaços, a opção por uma ou outra forma de representação de trajetórias está mais dependente do objetivo da representação e dos dados utilizados, do que propriamente de uma ou outra representação ter maior ou menor complexidade ou ser visualmente mais ou menos ilustrativa.

Andrienko (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013) apresenta uma representação diferente de movimentação humana e ocupação de espaço para um indivíduo ao longo de um dia. Neste caso agregam na mesma representação informação que permite ilustrar temporal e geograficamente as estadias de um indivíduo, assim como as trajetórias realizadas entre os vários locais visitados pelo indivíduo. Na Figura 2.14 são assim apresentadas as trajetórias que o indivíduo

realizou num dia (a verde), os locais visitados por este (pontos azuis) e uma representação do tempo de estadia em cada um dos locais (gráfico a amarelo). Segundo os autores, o gráfico maior (amarelo) corresponde possivelmente à habitação do indivíduo, enquanto o gráfico na parte inferior da imagem corresponde possivelmente ao local de trabalho. O gráfico apresentado no meio da imagem é possivelmente uma paragem diária do indivíduo com duas ocorrências no dia, uma de manhã e outra ao final da tarde que, segundo os autores, devido à curta duração dessa paragem, pode ser por exemplo uma paragem para deixar ou levar um familiar.

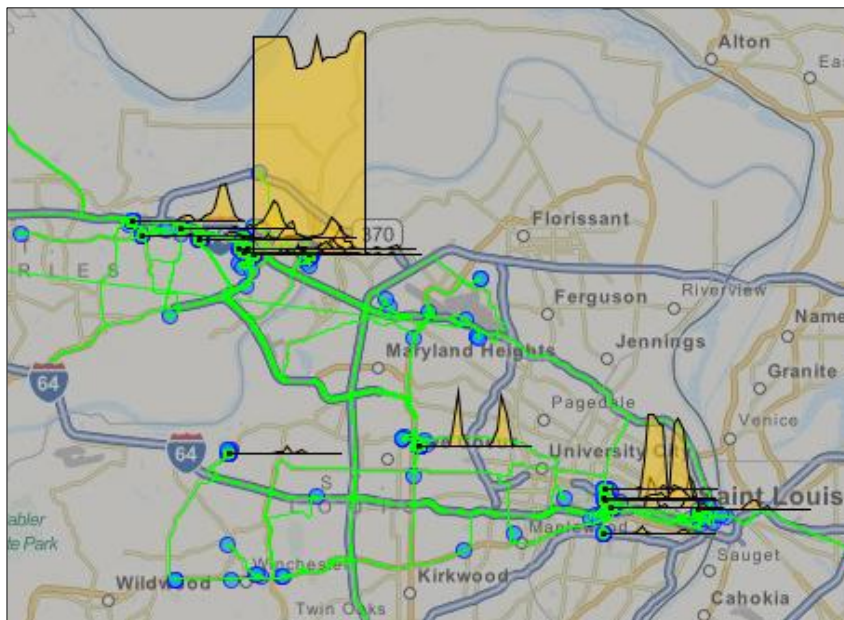


Figura 2.14 – Representação de locais visitados e trajetórias (G. Andrienko, Andrienko, Fuchs, et al., 2013)

Ainda no mesmo trabalho da Figura 2.14, os autores (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013) apresentam uma forma ainda mais complexa de representação de trajetórias realizadas por um indivíduo mas em diferentes dias, onde cada camada do gráfico corresponde a um desses dias (Figura 2.15). Este tipo de representação permite avaliar comportamentos diferentes de acordo com os dias da semana.

Embora não seja de todo uma representação de fácil leitura, este tipo de representação, denominada de **cubo espaço-tempo**, está direcionada para uma análise de possíveis alterações comportamentais ao longo do tempo, onde o referencial tempo se encontra representado no eixo vertical.

No exemplo da figura seguinte constata-se que nos primeiros 5 dias, quer as trajetórias, quer os locais visitados são bastante semelhantes mantendo assim o indivíduo um comportamento idêntico,

acabando por o mesmo ser totalmente distinto no sexto dia (topo da figura) onde se verifica uma alteração dos locais visitados e por consequência nas trajetórias realizadas.

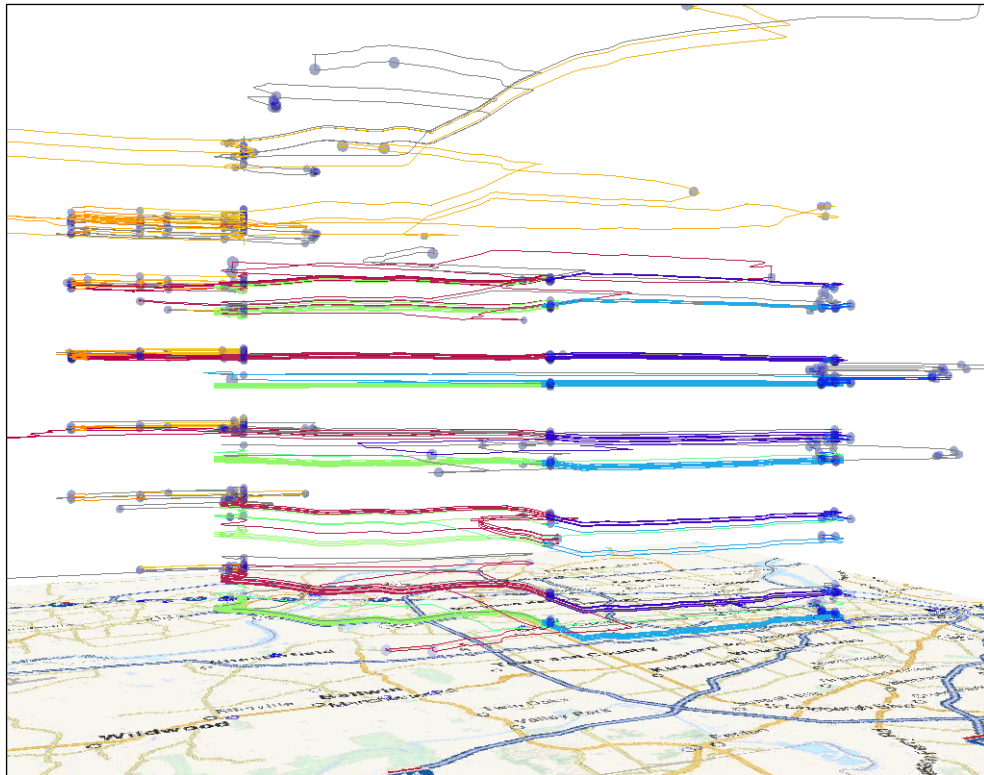


Figura 2.15 – Representação espaçotemporal de trajetórias e locais (G. Andrienko, Andrienko, Fuchs, et al., 2013)

A utilização de imagens estáticas para a representação de trajetórias e ocupação de espaço são importantes para a compreensão de determinados momentos temporais num espaço urbano. Para análises comparativas, por exemplo: períodos diurnos *versus* períodos noturnos ou dias da semana *versus* fins de semana; este tipo de visualização é uma grande mais-valia. No entanto, pelo facto de serem visualizações estáticas, perdemos a noção de atividade, de movimento, da vida que existe na própria cidade.

O recurso a **vídeos** para a visualização da mobilidade humana visa ilustrar as transformações que ocorrem, por exemplo, ao longo de um dia num espaço urbano, que de outra forma não seriam transmitidas pelos *snapshots* ou *footprints*. Os *footprints* pelo facto de serem acumulados de um período não poderiam pela sua natureza ilustrar as transformações ao longo do tempo, quanto muito ilustrariam sim as transformações mais significativas durante o período de análise. Vejamos por exemplo o caso de *footprints* de trajetórias (que são mais ilustrativas da mobilidade que as análises de ocupação de espaço), neste caso, as movimentações ou transformações realçadas são referentes a

todo o período de análise, não sendo possível observar apenas as transformações de um segmento de tempo específico. No caso dos *snapshots*, mesmo tendo vários ao longo do tempo e com curtos intervalos de tempo de duração (por exemplo uma hora), nunca se aproximam da natural continuidade temporal que permita representar as movimentações e transformações existentes num espaço urbano. Para termos o máximo de informação sobre as transformações existentes numa cidade seria necessário ter *snapshots* extraídos com uma periodicidade ao nível dos segundos, o que muito se assemelha a termos um vídeo construído pela sequência temporal de *snapshots*. Esta é uma das formas mais comuns de produzir um vídeo ilustrativo dos movimentos humanos, recorrendo à produção de *snapshots* e à sua apresentação sequencial.

Uma das ferramentas usadas para o desenvolvimento de projetos e trabalhos exploratórios na visualização de dados espaçotemporais recorre à metáfora das películas dos filmes, onde o efeito de movimento é originado pela rápida troca de imagens estáticas. A ferramenta de desenvolvimento *Processing*⁴ tem na sua sintaxe dois blocos que demonstram bem a ideia de produção de animações e vídeos com base na imagem do momento (método *draw*) e na apresentação sequencial dessas imagens (método *setup*).

```
void setup (){  
    size(640,360);  
    frameRate(30);  
}  
void draw(){  
    background(0);  
}
```

Figura 2.16 – Código exemplo de um projeto em *Processing*

Como podemos ver no exemplo de código anterior, a animação ou vídeo é feita com imagens geradas a cada 1/30 segundos e a imagem será composta com as características adicionadas no método *draw*.

⁴ <https://processing.org/> (visitado em Agosto de 2016)

Um dos primeiros vídeos/animação que maior visibilidade teve, desenvolvido em *Processing* por Aaron Koblin⁵ em 2006, apresenta uma visualização de um dia de trajetórias (embora de forma geograficamente pouco rigorosa) realizadas por aviões comerciais de, para e dentro dos Estados Unidos da América. A Figura 2.17 ilustra um *snapshot* extraído do vídeo disponível online⁶.



Figura 2.17 – Voos comerciais realizados num dia nos EUA

No vídeo anterior é possível visualizar quer os períodos do dia em que o tráfego aéreo é mais intenso, quer os momentos em que este mesmo tráfego diminui significativamente. Como esta visualização tem por base a representação das trajetórias dos aviões, conseguimos ainda extrair informação relativamente aos aeroportos com maior afluência tanto ao nível das partidas como das chegadas.

A grande vantagem dos vídeos relativamente aos *snapshots* é conseguirmos extrair a mesma informação destes se pararmos o vídeo em pontos-chave (por exemplo, de hora a hora) e em adição temos ainda a noção de como a mobilidade evolui ao longo do tempo e do espaço.

⁵ <http://www.aaronkoblin.com/> (visitado em Agosto de 2016)

⁶ Disponível em: http://www.aaronkoblin.com/work/flightpatterns/FPWeb_Final_3.mov (visitado em Agosto de 2016)

Uma das referências ao nível da visualização de mobilidade com recurso aos vídeos é o *MIT SENSEable City Lab*⁷. Este grupo de investigação tem um conjunto alargado de vídeos de mobilidade humana e ocupação de espaço. Um dos trabalhos iniciais com maior destaque refere-se ao projeto *Real Time Rome*. Num dos vídeos produzidos neste projeto⁸ é feita uma visualização da ocupação do espaço com base na informação da utilização da rede celular para dois eventos em particular: o campeonato do mundo de futebol e um concerto de uma estrela da música. Com base na análise do vídeo é perceptível quais as zonas da cidade com maior concentração de indivíduos de acordo com o evento (localização deste e tempo em que o mesmo ocorre). Embora as conclusões sobre a análise possam parecer óbvias, isto é, é perfeitamente normal que de acordo com este tipo de eventos exista uma correlação com a ocupação do espaço, a inovação e o que distinguiu esta visualização das existentes até então foi permitir sentir de que forma a cidade (em particular os seus indivíduos) respondem à ocorrência de dois eventos importantíssimos. Esta visualização permite verificar como é a ocupação do espaço antes, durante e após os eventos e com a comparação de dois eventos distintos, quais os locais de relevo tendo em consideração o tipo de evento que está a ocorrer. No âmbito do projeto *Wiki City Rome* foi disponibilizada⁹ uma visualização em vídeo, tendo por base o projeto anterior, que uma vez mais ilustra como a ocupação da cidade evolui ao longo do tempo de acordo com os eventos que estão a acontecer nessa cidade (neste caso de menor importância relativamente aos dois do projeto anterior) e com a taxa de utilização de uma rede celular disponível. Com base na informação dos autocarros é ainda possível inferir (com base no rasto representado no vídeo) quais as artérias com maior ou menor fluxo de trânsito.

Nos dois projetos referidos anteriormente o foco foi mais a forma como a ocupação do espaço evolui com o tempo. Mas, como vimos na Figura 2.17, os vídeos também podem constituir uma forma de visualização interessante para a representação de trajetórias. No projeto *City Motion* no âmbito do programa *MIT Portugal*, um dos resultados produzidos em vídeo¹⁰ com maior destaque faz a representação de trajetórias realizadas pelos táxis na cidade de Lisboa (com base em informação de

⁷ Página do grupo de investigação: <http://senseable.mit.edu/> (visitado em Agosto de 2016)

⁸ Disponível em: <https://www.youtube.com/watch?v=SgbCgA3O06s> (visitado em Agosto de 2016)

⁹ Disponível em: <https://www.youtube.com/watch?v=4DnRm3n6sco> (visitado em Agosto de 2016)

¹⁰ Disponível em: <https://www.youtube.com/watch?v=ODR5M564-6k> (visitado em Agosto de 2016)

GPS: localização e velocidade) e com base nessa representação é possível ao longo do tempo perceber quais as áreas da cidade com maior fluxo de trânsito e como o fluxo ou congestionamento aumenta e diminui com o tempo.

A utilização do vídeo como visualização do movimento humano permite ter ganhos relativamente às imagens estáticas: analisar de forma rápida, momentos temporais diferentes e perceber de que forma o espaço se transforma ou se adapta à mobilidade e sua ocupação por parte dos indivíduos. No entanto existe uma rigidez comum quer nos vídeos quer nas imagens estáticas. Após a sua produção não é possível fazer qualquer manipulação dos dados de forma a alterarmos os parâmetros da nossa análise. Até uma simples tarefa de aproximar ou afastar o mapa para revelar mais ou menos detalhe (e muitas das visualizações são baseadas em representação geográfica) é totalmente impossível de realizar.

Uma forma de minimizar esta rigidez é adicionar interatividade às visualizações. Nesse sentido temos uma diferente técnica de visualização que poderemos apelidar de **demonstrações interativas**.

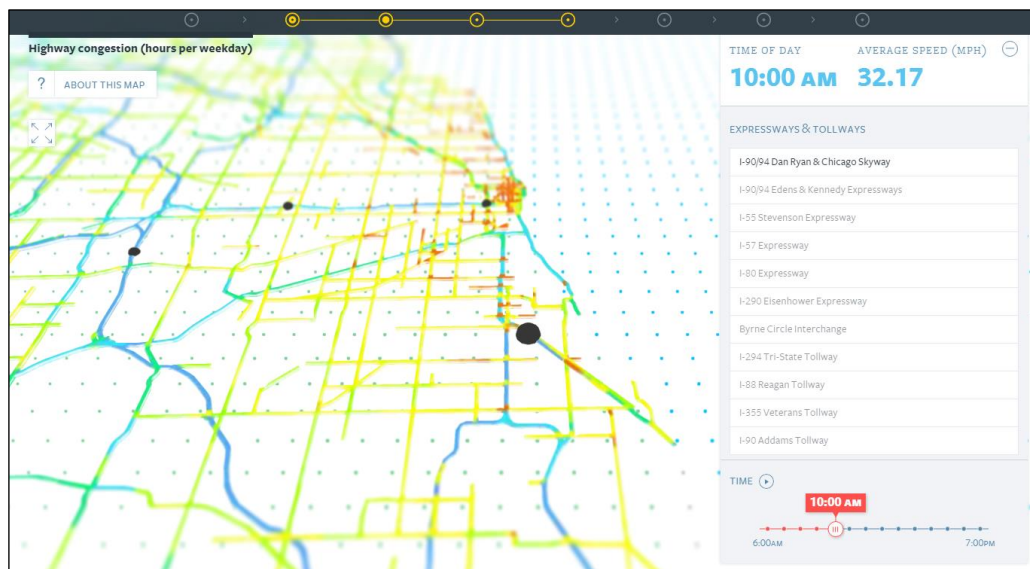


Figura 2.18 – Congestionamento na cidade de Illinois (Planning, 2014)

As demonstrações interativas são assim uma visualização que pode ou não ser animada (no tempo) semelhante ao vídeo ou que apenas permite a interação com os objetos de análise, com a vantagem de permitir o controlo ou do próprio tempo da visualização ou das componentes que estamos a analisar.

Um desses exemplos¹¹ está ilustrado na Figura 2.18. Nesta demonstração interativa é possível controlar a via urbana que pretendemos estudar ao nível do congestionamento de tráfego e manipular o momento temporal. Desta forma conseguimos uma análise mais detalhada do que a pura análise de um vídeo com a sequência temporal de eventos. Neste caso é ainda possível manipular o objeto que pretendemos analisar e o momento temporal da análise. As duas figuras seguintes demonstram como a manipulação que fazemos do tempo e espaço nestas demonstrações interativas permitem extrair conclusões interessantes que possivelmente de outras formas as mesmas seriam e permaneceriam ofuscadas pelos dados¹². Enquanto na Figura 2.19 verificamos que a população de uma cidade do Estado de Nevada sai de casa para o trabalho maioritariamente antes das 5h da manhã, na Figura 2.20 uma outra cidade, neste caso do Estado de Dakota do Sul, tem hábitos diferentes e maioritariamente a população sai para o trabalho entre as 7h30m e as 8h da manhã.

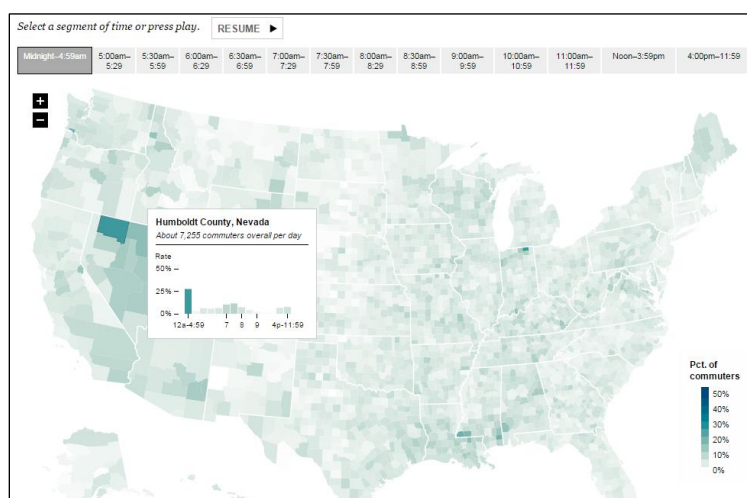


Figura 2.19 – Saída para o trabalho, caso 1 (Yau, 2014)

¹¹ Disponível em: <http://www.cmap.illinois.gov/mobility/explore#/topic/roads/congestion> (visitado em Agosto de 2016)

¹² Disponível em: <http://flowingdata.com/2015/02/04/when-do-americans-leave-for-work/> (visitado em Agosto de 2016)

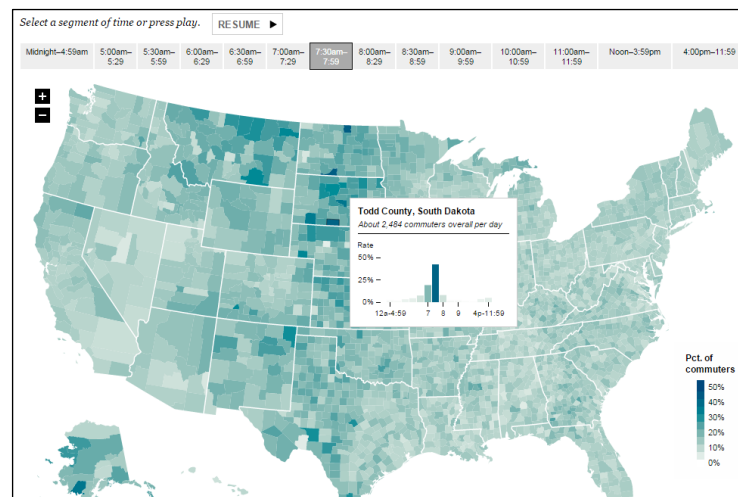


Figura 2.20 – Saída para o trabalho, caso 2 (Yau, 2014)

Um outro exemplo de uma demonstração interativa com múltiplas hipóteses de combinações é o projeto *ManyCities*¹³ (Grauwin *et al.*, 2014) (Figura 2.21).

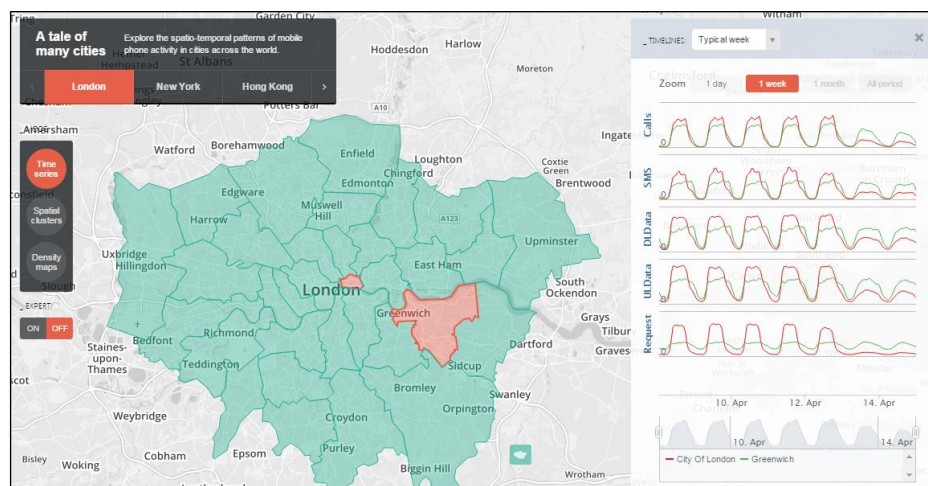


Figura 2.21 – Projeto ManyCities (Grauwin et al., 2014)

Nesta demonstração interativa é possível testar as nossas teorias em três diferentes cidades, comparando diferentes comportamentos de acordo com diferentes zonas da cidade, através da seleção de características e manipulação dos dados disponibilizados.

¹³ Disponível em: <http://www.manycities.org/> (visitado em Agosto de 2016)

No entanto nem sempre acontece termos demonstrações interativas onde seja possível manipularmos em simultâneo os objetos de análise e o tempo. Estas podem ser mais simples e contemplar apenas uma das duas vertentes referidas anteriormente. Podem, por exemplo, apenas permitir a seleção do objeto que pretendemos analisar sem ter a componente temporal.



Figura 2.22 – Exemplo da seleção dos objetos de análise (Cabalero e Belmonte, 2013)

No caso da Figura 2.22 é possível apenas selecionar quais os operadores de aviação comercial que pretendemos visualizar, enquanto na Figura 2.23 a possibilidade de seleção está restrita às várias regiões administrativas existentes nos Estados Unidos.

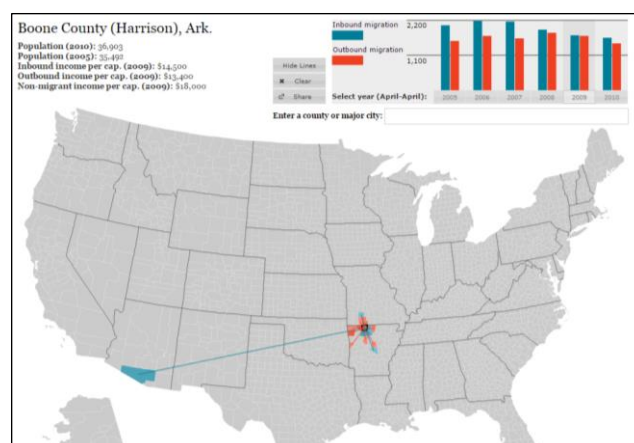


Figura 2.23 – Migração ao longo dos Estados Unidos da América (Bruner, 2012)

Ou então permitir que se possa manipular apenas a componente temporal, como acontece na Figura 2.24 em que podemos visualizar a informação de acordo com os dias da semana e intervalos de tempo pré-definidos.

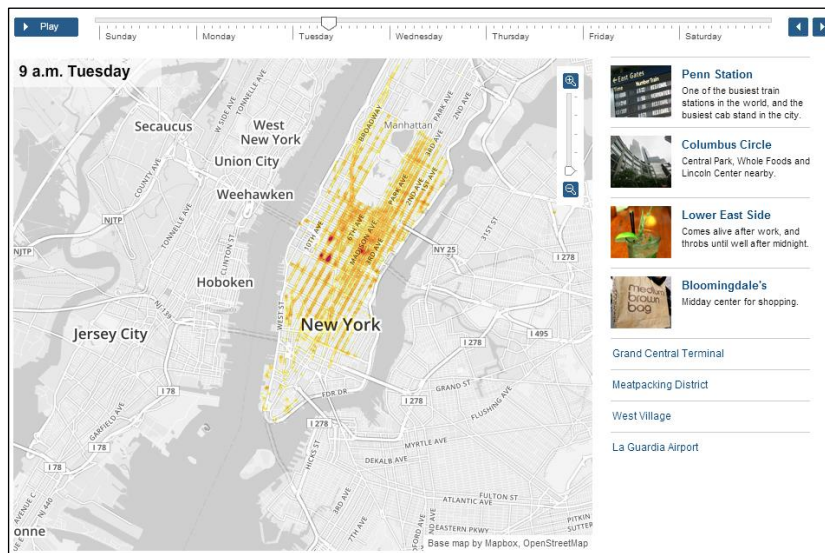


Figura 2.24 – Fluxo de táxis ao longo do tempo (Bloch, Fessenden e Carter, 2009)

A Figura 2.25 apresenta um outro exemplo da manipulação apenas da componente temporal, em particular, uma perspectiva da utilização do metro ao longo de um dia.

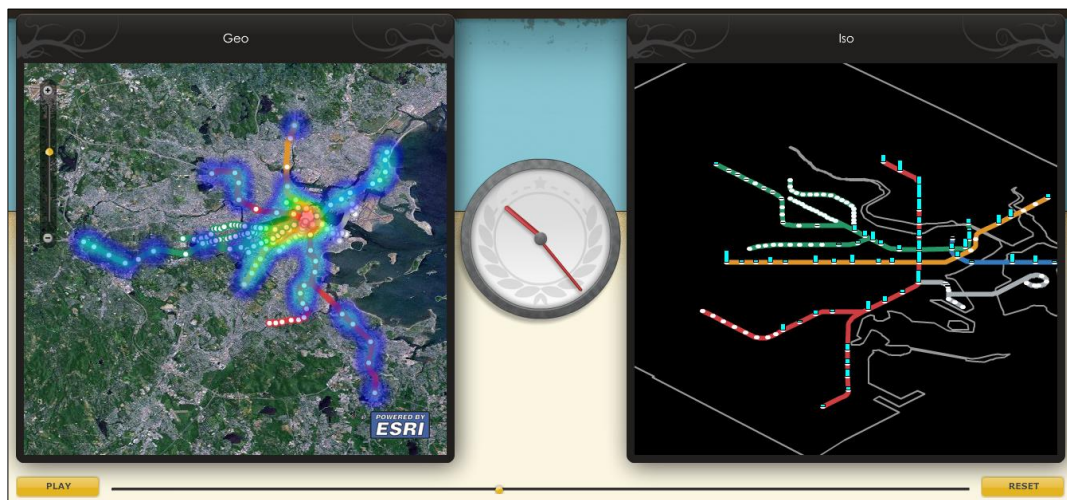


Figura 2.25 – Taxa de utilização do metro ao longo do tempo (ESRI, 2016)

A visualização do movimento humano, quer através da representação de trajetórias dos movimentos realizados pelos artefactos, quer através da representação da ocupação do espaço urbano, pode atualmente ser realizada de diversas formas de acordo não só com os objetivos em si da representação, mas também dependentes da informação disponível para a elaboração da visualização.

Não se pode afirmar que exista uma visualização diferenciadora e com qualidades superiores às restantes, mas sim que, de acordo com os requisitos da visualização e das características da

informação, existem visualizações mais adequadas que outras, fomentando assim a diversidade de visualizações apresentadas ao longo deste capítulo.

2.4. Observação do movimento humano

Estudar as dinâmicas das cidades pressupõe a existência de informação sobre as movimentações dos artefactos no espaço urbano. Logo, a aquisição de dados que diretamente ou indiretamente reflitam essas movimentações é um fator importante e que condiciona em muito o estudo que se pretende realizar.

Relativamente à forma como são adquiridos os dados usados nos estudos de mobilidade humana, a literatura apresenta um conjunto alargado de possibilidades, cada uma delas com as suas vantagens e desvantagens.

Uma das formas mais antigas e que ainda se tem usado (Jiang, Ferreira e Gonzalez, 2012) para a recolha de dados passa pela realização de questionários/inquéritos à população que se pretende estudar. Embora seja uma metodologia com um custo elevado tendo em consideração a necessidade de recrutar quem faça os inquéritos/entrevistas aos cidadãos; ser um processo moroso uma vez que em alguns casos esses inquéritos são manuais e é necessário realizar todo um processo de inserção dos dados para posterior tratamento; na perspetiva do estudo das dinâmicas existem ainda assim um conjunto alargado de desvantagens. Primeira delas, a duração da validade dos resultados que se obtém, não porque os resultados em si sejam menos válidos, mas porque os dados que são recolhidos são estáticos e como tal com o avançar do tempo torna-se cada vez mais inadequada a sua utilização. Imaginemos que temos um conjunto de respostas a um questionário de mobilidade que foi realizado há 10 anos atrás: muito possivelmente, com as transformações que acontecem diariamente nos espaços urbanos, a informação recolhida já não reflete a situação atual e como tal as conclusões que possamos vir a extrair podem estar ligeiramente desalinhadas com a realidade. Esta conclusão está relacionada com a segunda grande desvantagem, a resolução temporal. Os inquéritos são realizados com uma frequência temporal alargada, pelo que não é fácil obter dados com períodos de amostragem muito curtos e assim dificilmente se consegue estudar as variações ao longo de meses. Terceira grande desvantagem, a resolução espacial. Dificilmente nos inquéritos conseguimos ter uma resolução espacial na ordem dos metros, contrariamente ao que conseguimos com um recetor GPS. Quarta desvantagem, a abrangência espacial. Realizar inquéritos para grandes áreas geográficas torna o seu

custo elevadíssimo não sendo por isso aceitável a sua realização. E poderíamos continuar a descrever mais algumas das desvantagens desta abordagem para a observação do movimento humano, no entanto as que foram apresentadas anteriormente são suficientes para justificar o porquê de em vários trabalhos de mobilidade terem surgido alternativas à recolha manual e à observação estática da realidade, optando-se assim por abordagens que permitam automatizar a recolha de dados, simplificando ao máximo este processo, e adicionando a característica dinâmica aos próprios dados, isto é, permitir ter o mais possível dados recentes. Estes dados não têm que ser recolhidos e analisados em tempo real, não tem que ser um estudo ao instante, mas que permitam que a distância temporal (entre a recolha e a análise) seja relativamente curta garantindo-se desta forma uma maior adequação dos dados e consequentes conclusões, face ao momento atual.

Uma abordagem muito usada por vários autores recorre à utilização de dados gerados pelas **redes sociais** e a partir destes inferirem potenciais movimentações no espaço de acordo com a informação publicada ao longo do tempo. A utilização das publicações em redes sociais como base para a observação de fenómenos é atualmente muito usada. São vários os exemplos e nas mais variadas áreas, como por exemplo, marketing e vendas (Huberman, Romero e Wu, 2009), onde estas são usadas, por exemplo, para extrair a perceção que os compradores têm de uma marca (Jansen *et al.*, 2009).

Relativamente ao estudo da mobilidade humana e consequente dinâmica das cidades, uma das redes sociais utilizada como “fornecedora” de dados que permite inferir a mobilidade é o **Twitter**. Como Sakaki refere (Sakaki, Okazaki e Matsuo, 2010) este tipo de **sensor social** embora não esteja ativo em permanência, pois a publicação de *tweets* está dependente da ocorrência de um qualquer evento (mesmo que irrelevante para a comunidade) e que este sensor possa estar inoperacional durante os períodos em que os utilizadores estão *offline*, o facto de existirem milhões de potenciais sensores a publicar *tweets* pelo mundo, em conjugação com a possibilidade de um *tweet* poder ser descrito num referencial temporal e geográfico, torna esta rede social como uma forte candidata a utilização da sua informação para a caracterização da mobilidade humana.

Sakaki alerta para uma questão importante relativamente à representação geográfica nas redes sociais: muitas vezes a informação da localização refere-se ao registo do utilizador e não ao local onde foi realizada a publicação na rede social.

Cada *tweet* permite extrair um mínimo de informação (Tabela 2.1), contextualizando o local e o instante em que o *tweet* foi publicado na rede social e o texto publicado no respetivo *tweet*.

id_user	time	latitude	longitude	text
---------	------	----------	-----------	------

Tabela 2.1 – Informação mínima extraída de um *tweet*

Nem sempre estes cinco atributos contêm informação válida, podendo existir por exemplo *tweets* sem localização, ou existir informação que, dependendo da análise de mobilidade que se pretende, está em excesso, como por exemplo o texto do *tweet*, e como tal não ser utilizada.

A recolha de dados oriundos do *Twitter* passa normalmente pela utilização da própria *API* desta rede social. Nesse sentido poderão ser recolhidos dois conjuntos distintos: dados georreferenciados (Frias-Martinez e Frias-Martinez, 2014) onde se dá importância ao momento temporal e posição geográfica da publicação, ou aos conteúdos textuais das publicações (G. Andrienko, Andrienko, Bosch, *et al.*, 2013), onde se faz a interpretação do texto publicado.

Como MacEachren (MacEachren e Robinson, 2011) refere, a utilização dos dados recolhidos do *Twitter* pode ser feita das seguintes formas, partindo da mais básica e aumentando a sua complexidade: simples representação dos *tweets* num mapa; contagem e/ou agregação dos *tweets* por local ou área geográfica; por último, representação espaçotemporal dos *tweets*.

Uma outra rede social cujos dados também são utilizados para a caracterização da ocupação do espaço e/ou mobilidade humana é o **Facebook**. Chang (Chang, Sun e Ave, 2010) realizaram um estudo usando dados recolhidos a partir do *Facebook Places* (*check-ins* e Pontos de Interesse) e com base apenas nestes dados estudaram a cidade de São Francisco. Com este trabalho não só conseguiram identificar “o quando” e “o onde” os indivíduos se encontram fisicamente, e com base nessa informação construir *heatmaps* de ocupação de espaço, como também desenvolveram um modelo que permite prever com bastante precisão os próximos *check-ins* com base particularmente nos *check-ins* realizados anteriormente e nos *check-ins* realizados pelos amigos.

Além das duas redes sociais referidas até aqui existem outras duas que pelas suas características também podem e são usadas para inferir questões associadas à mobilidade humana, são elas o **Foursquare** e o **Instagram**. Ribeiro (Ribeiro *et al.*, 2014) recorre a dados destas duas redes sociais (*check-ins*) e correlacionam os mesmos com os dados de trânsito obtidos do *Bing Maps* (os dados não são acedidos de forma simples, mas através da captura de imagem e tratamento das mesmas no sentido de obter informação sobre o fluxo de trânsito). A correlação entre os resultados obtidos através das duas redes sociais e os resultados obtidos pelo processamento das imagens dos mapas é significativa não se verificando uma diferença relevante, apenas se notando um desfasamento temporal com um atraso que poderá chegar aos 30 minutos. Este desfasamento temporal seria

expectável uma vez que depende sempre do momento em que são feitas as publicações nas redes sociais (normalmente feitas nos locais de *check-in* e não durante o período em que os indivíduos se encontram em trânsito).

As redes sociais não têm como principal objetivo a localização de pessoas no espaço e no tempo. Estes dados são produzidos de forma indireta decorrente da atividade de publicação nas redes. São assim **Sistemas Indiretos de Observação**. Dentro deste grupo de sistemas podemos ainda incluir outros tipos de sensores que mais uma vez não são diretamente vocacionados para a análise de mobilidade, mas pelos dados que produzem têm condições para que os mesmos possam ser usados para inferir aspetos relevantes da mobilidade humana. Incluem-se nesta categoria os dados obtidos a partir da utilização de infraestruturas de telecomunicações, como é o caso das **redes celulares**.

As redes celulares têm como principal propósito disponibilizar serviços de voz e dados aos seus utilizadores pelo que o foco não passa por estudar a mobilidade humana. No entanto existem trabalhos que partindo dos dados gerados pela utilização destas redes conseguem extrair conclusões relevantes para caracterizar movimentos e/ou ocupação de espaço. Nesta área não podemos falar de redes celulares sem falar do trabalho de Reades (Reades *et al.*, 2007). Possivelmente dos primeiros trabalhos a usar dados de redes celulares para este fim, os seus autores conseguiram demonstrar a evolução da ocupação do espaço ao longo do tempo para 6 espaços com características distintas, permitindo assim ilustrar as variações entre os espaços e dentro de cada espaço.

González (González, Hidalgo e Barabási, 2008) com recurso também a dados recolhidos da atividade natural das redes celulares extraíram conclusões interessantes quer ao nível da utilização ou não deste tipo de dados quer referentes à mobilidade humana. Relativamente ao primeiro aspeto verificaram que o facto de transportarmos connosco os telemóveis durante as nossas rotinas diárias torna os telemóveis, e consequentemente os dados gerados diretamente ou indiretamente por estes, um dos melhores *proxies* para capturar as trajetórias humanas. Ao nível da mobilidade, os autores concluíram que existe uma forte tendência para os indivíduos estudados retornarem a locais previamente visitados e que estes indivíduos passam a maior parte do seu tempo num número reduzido de locais, como verificado anteriormente na seção 2.2.

Andrienko (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013) conclui, com base também em dados da utilização das redes celulares, que a informação extraída não permite representar pequenas mudanças espaciais e como tal pequenas estadias. Esta limitação, quando usados dados das redes

celulares isoladamente, pode camuflar visitas de curta duração a locais, mas que pelo carácter repetitivo de visita poderão ser locais relevantes para os indivíduos.

Relativamente aos dados recolhidos a partir da utilização das redes celulares, estes podem medir a taxa de utilização das estações base, usando como unidade de medida o Erlang¹⁴ e desta forma ter um indicador da densidade de indivíduos que se encontram nas proximidades de uma determinada antena de rede móvel (Reades *et al.*, 2007), ou podem descrever o instante em que existe alguma atividade nas comunicações (iniciar ou receber uma chamada e enviar ou receber uma mensagem de texto) e a localização geográfica da antena que está associada ao equipamento (González, Hidalgo e Barabási, 2008). A representação dos dados recolhidos a partir da utilização das redes celulares poderá ser próxima da Tabela 2.2:

time	cellid	erlang	latitude	longitude
------	--------	--------	----------	-----------

Tabela 2.2 – Potencial informação obtida das redes celulares

onde potencialmente se poderá saber a taxa de utilização (*erlang*) num instante *time*, de uma determinada *cellid*, cuja localização é dada pela *latitude* e *longitude*. Ou como González descreveu, uma representação mais próxima da Tabela 2.3:

time	cellid	latitude	longitude	mobile_user
------	--------	----------	-----------	-------------

Tabela 2.3 – Outra potencial representação dos dados das redes celulares

onde para cada *mobile_user* anonimizado se consegue identificar o instante temporal *time* e a localização da *cellid* em que ocorreu um evento de comunicação.

Uma das grandes dificuldades dos dados das redes celulares é a sua disponibilidade. Como são dados operacionais obtidos da atividade dos operadores, a sua disponibilização é dificultada por questões de privacidade e o acesso aos mesmos por parte da comunidade científica é excessivamente complexa. Principalmente por esse motivo os investigadores exploram outras potenciais fontes de dados, como é o caso, por exemplo, das **redes Wi-Fi**.

¹⁴ Como unidade de medida da intensidade de tráfego das redes de comunicação, 1 Erlang corresponde a uma pessoa a falar ao telemóvel durante uma hora ou duas pessoas a falarem por meia hora (Reades *et al.*, 2007)

Sevtsuk (Sevtsuk e Ratti, 2005) num trabalho realizado com dados da rede Wi-Fi do *campus* do MIT apresentam resultados de ocupação do espaço em diferentes momentos temporais com base no número de utilizadores a utilizarem esta rede sem fios. A Tabela 2.4 ilustra parte da informação recolhida para a realização do trabalho, em particular o instante temporal (*time*) em que a recolha foi realizada, a informação relativa ao ponto de acesso Wi-Fi (*AP*) que está a ser observado e o número de utilizadores ativos (*usage*) ligados no momento a esse AP.

time	AP	usage
-------------	-----------	--------------

Tabela 2.4 – Informação parcial recolhida sobre a utilização da rede Wi-Fi

Relacionando a informação da utilização da rede com a categorização dos espaços, os autores conseguem ainda caracterizar no tempo os locais com maior ou menor taxa de ocupação. Um dos exemplos referidos é relativo à cafetaria onde, entre as 7h e as 11h da manhã, se verifica a maior taxa de utilização da rede.

No trabalho de Sevtsuk os dados recolhidos referem-se à utilização da rede Wi-Fi, muito semelhante à informação usada por Reades relativa à utilização das redes móveis celulares. Embora o trabalho de Sevtsuk não indique de forma explícita a representação da localização dos pontos de acesso, tudo leva a crer que a mesma seja simbólica uma vez que realizaram a associação da localização dos vários pontos de acesso da rede Wi-Fi com as diferentes divisões (salas, cafetaria, entre outros) existentes nos diferentes edifícios do *campus* do MIT.

Como a mobilidade humana está muito relacionada com a forma ou meio como nos deslocamos na cidade, vários autores exploram assim dados oriundos de diferentes meios de transporte, em particular transportes públicos, na tentativa de extrair conclusões relativamente à mobilidade. Foi o que Liu (Liu, Biderman e Ratti, 2009) fez utilizando **dados de bilhética** baseados em cartões de viagem que podem ser usados em **autocarro** e **metro**. Os dados recolhidos seguem a mesma estrutura independentemente do modo de transporte:

Timestamp	CardId	Tradetype	Terminalid	Trademoney
------------------	---------------	------------------	-------------------	-------------------

Tabela 2.5 – Dados de bilhética (Liu, Biderman e Ratti, 2009)

Os dados relativos ao modo de transporte autocarro que são registados têm a ver com a hora de entrada no autocarro (*timestamp*), o identificador da paragem (*terminalid*) e a tarifa da viagem (*trademoney*). No caso dos dados relativos ao metro são registados os locais e tempo de entrada e

saída das paragens (*terminalid*). Neste último caso, cada viagem terá dois registos contrariamente aos dados relativos ao autocarro.

Com base nos dados recolhidos verificaram a existência de padrões temporais na utilização destes dois modos de transporte público. No caso das viagens de autocarro verificaram: para os dias de semana um pico no início do dia e outro no final do dia; ao sábado estes picos mantêm-se, possivelmente porque grande parte da população alvo trabalha também nesse dia; e um comportamento diferente, como seria de esperar, ao domingo. Comparativamente com os padrões detetados nos dados referentes ao metro estes mantêm-se, apenas se verifica um desfasamento temporal nos picos. Como os dados associados aos autocarros não têm registo da saída, contrariamente ao que acontece com os do metro, não é possível perceber a origem e destino dos movimentos. Assim neste trabalho foi apenas estudado de forma espaçotemporal os dados relativos à utilização do metro. Foi possível assim revelar entre que zonas da cidade os indivíduos se deslocam com diferentes padrões ao longo da semana, com natural diferença para o domingo (neste caso particular o sábado é também um dia de trabalho para a maioria da população). Bhattacharya (Bhattacharya *et al.*, 2013) usando apenas dados de bilhética de autocarros da cidade de Lisboa, pretendeu caracterizar os padrões de viagem dos utilizadores de autocarro de forma a explicar a dinâmica da cidade. Neste trabalho os autores com a informação de bilhética criaram um modelo baseado em informação contextual onde por exemplo além da informação de bilhética, o estado do tempo era também um dos critérios usados para caracterizar a utilização deste meio de transporte. Utilizando o mesmo conjunto de dados de Lisboa, Foell (Foell *et al.*, 2014) conclui ainda que as previsões com maior grau de certeza são as que consideram para o seu cálculo: o histórico pessoal das viagens realizadas pelo indivíduo e os padrões de mobilidade dos restantes indivíduos. Demonstram assim, em certa medida, o carácter repetitivo da mobilidade deste tipo de utilizadores de transporte público. Anteriormente, com base na informação de autocarros, mas neste caso com dados de GPS e não de bilhética, Pinelli (Pinelli *et al.*, 2009) desenvolveu um algoritmo de extração das linhas/trajetos realizados por estes e respetivas paragens usando para tal a informação de localização **GPS** de cada autocarro, composta por uma sequência de registos com informação do instante temporal e localização, semelhante à estrutura apresentada na Tabela 2.6.

timestamp	latitude	Longitude
-----------	----------	-----------

Tabela 2.6 – Estrutura de dados GPS

Esta é a estrutura mínima associada à informação extraída de um recetor de GPS, suficiente para caracterizar um artefacto no tempo e no espaço, podendo no entanto ser adiciona outra informação como por exemplo: velocidade, orientação, entre outros.

Atualmente a tecnologia GPS não serve apenas como ferramenta de orientação e navegação, mas também como ferramenta para capturar as viagens realizadas (Van der Spek *et al.*, 2009). Com a sua divulgação cada vez mais ampla: smartphones, câmaras fotográficas, sistemas de navegação dos veículos, entre outros; cada vez mais os movimentos que realizamos (no exterior) podem ser recolhidos com recurso a esta tecnologia. De forma a validarem a estrutura desenvolvida, Andrienko (Andrienko *et al.*, 2010) utilizou dados de GPS circunscritos à cidade de Milão com o objetivo de encontrarem padrões temporais no trânsito. Usando apenas esta tipologia de dados conseguiram demonstrar o que seria do conhecimento do senso comum, o trânsito é complicado durante as manhãs dos dias da semana com repetição ao final do dia, durante a noite existe pouca afluência de tráfego e ao fim de semana o padrão é distinto comparativamente aos dias da semana. Estas conclusões confirmam assim o que seria esperado. Como resultado deste trabalho foi possível identificar os padrões de trânsito ao longo do dia e consequentes variações, assim como, identificar espacialmente quais as zonas da cidade com maior ou menor concentração de tráfego. Dando continuidade ao seu trabalho, Andrienko (Andrienko *et al.*, 2011) recorre novamente a dados de GPS provenientes de cerca de 8 mil carros, referentes a apenas um dia. Com base nestes dados extraíram apenas as ocorrências em que a velocidade instantânea era baixa de forma a serem calculados *clusters* baseados na localização e no tempo (informação proveniente também dos dados de GPS). Estes *clusters* permitiram identificar zonas de congestionamento e períodos temporais em que as mesmas ocorreram. Posteriormente, Andrienko (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013) recorre mais uma vez a *traces* de GPS neste caso para extrair padrões de mobilidade individual. Como referem nesse trabalho, se a recolha de dados de GPS for feita durante um longo período temporal, é possível identificar três tipos de comportamentos: rotineiros, esporádicos ou fora do habitual. A utilização dos dados de GPS, além de permitirem a reconstrução dos movimentos realizados, permite também a extração dos locais visitados. Zheng (Zheng *et al.*, 2009) utiliza esta informação para extrair pontos de interesse visitados pelos indivíduos observados e as respetivas sequências de visita.

A grande dificuldade da utilização e recolha de dados através da tecnologia GPS prende-se principalmente com o consumo de energia. Mesmo com os recentes avanços ao nível da autonomia energética dos dispositivos, por exemplo dos smartphones, tal não é suficiente para garantir um dia

inteiro de recolha de dados GPS. No entanto existem cenários onde a recolha de dados GPS pode ser feita uma vez que esse entrave é mínimo. É o caso, como vimos anteriormente, da integração desta tecnologia nos autocarros ou, por exemplo, nos **táxis**.

Na literatura existem alguns exemplos onde *traces* GPS provenientes de táxis foram já utilizados para o estudo da mobilidade humana. Um desses exemplos é o trabalho realizado por Liu (Liu, Biderman e Ratti, 2009) que utilizaram no seu estudo dados gerados durante um mês por táxis equipados com GPS. Uma conclusão interessante deste estudo é o período temporal em que a utilização dos táxis é maior. Com base nos resultados os autores verificaram que o dia da semana de maior atividade dos táxis é o sábado, considerando assim que os táxis naquela cidade eram usados para viagens recreativas e/ou compras e não tanto para a deslocação para os locais de trabalho. Além desta conclusão foi possível ainda caracterizar a utilização do táxi de acordo com a hora do dia.

O trabalho anterior foca apenas a componente temporal na análise de mobilidade. No entanto como Ferreira (Ferreira, Poco e Vo, 2013) refere, as viagens realizadas pelos táxis são interessantes ao nível da análise de mobilidade porque os seus dados podem conter a componente temporal e geográfica. Neste trabalho os dados usados eram compostos pela informação (localização e tempo) relativa às ações de recolher e largar clientes deste meio de transporte, sendo assim a análise é feita com base na viagem. A adição de informação da viagem à informação comum do GPS permite assim definir de onde e para onde a mesma se realizou e em que momentos temporais. Desta forma foi possível aos autores identificar padrões no tempo e no espaço associados à mobilidade baseada nas viagens de táxi.

Até ao momento verificámos que os dados gerados por diferentes infraestruturas de transporte público podem com sucesso ser utilizados para caracterizar a mobilidade num espaço urbano. Mas as cidades estão cada vez mais preocupadas com as questões ambientais e congestionamento de tráfego, surgindo por isso, felizmente, um número crescente de cenários de partilha de meios de transporte ecológicos, como as **bicicletas**. Froehlich (Froehlich, Neumann e Oliver, 2009) desenvolveu um trabalho interessante na análise espaçotemporal da cidade de Barcelona tirando partido das infraestruturas existentes na cidade. Esta é uma tipologia de dados distinta de todas as outras referidas até ao momento, que utiliza o sistema de *bike sharing* da cidade como fonte de dados. Os dados gerados pela infraestrutura, em particular, relacionados com o *check-in* e *check-out* no levantamento e entrega das bicicletas nos vários parques disponíveis, permitiram aos autores estudar padrões de mobilidade. Não só conseguiram compreender a dinâmica da cidade ao longo do tempo e espaço

como conseguiram criar modelos preditivos da utilização das bicicletas e parques, sendo possível prever com 80% de certeza a evolução destes nas próximas duas horas. A estrutura dos dados recolhidos a partir da infraestrutura é a seguinte:

timestamp	station		available bicycles	parking slots
	latitude	longitude		

Tabela 2.7 – Dados obtidos da infraestrutura de *bike sharing*

Neste caso a novidade são os atributos *available bicycles* e *parking slots*, que indicam respetivamente quantas bicicletas e quantos lugares de estacionamento estão disponíveis na *station*.

Mas não são só as infraestruturas associadas às bicicletas que podem ser usadas como fonte de dados para a observação de movimentações dentro do espaço urbano. As **portagens**, embora tenham um objetivo específico, produzem dados que identificam as entradas e saídas de veículos ou a passagem destes por determinadas zonas, sendo por isso uma fonte de dados aliciante para a análise do movimento humano. Como refere Bennett (Bennett, Raab e Regan, 2003) esta forma de “vigilância” adiciona inteligência à gestão de tráfego.

Os tipos de dados obtidos pela utilização das infraestruturas até agora apresentados são adquiridos de forma indireta, resultantes da própria operação de uma infraestrutura criada para outros fins. No entanto estas infraestruturas não foram desenvolvidas com o objetivo final de recolherem dados de mobilidade. Por seu lado, o GPS, embora tenha como principal objetivo a orientação e navegação, têm pela sua natureza de funcionamento dados de mobilidade. A informação sobre localização (latitude e longitude), velocidade e tempo, disponibilizada por esta tecnologia tem as características necessárias para uma análise espaçotemporal da mobilidade de artefactos. Mas, além da rede de satélites que permite a navegação através de GPS, existem outras infraestruturas não geoestacionárias baseadas também em **satélites** cujo objetivo passa em certa medida pela recolha de dados que poderão diretamente ser usados para análise de mobilidade humana.

Ramachandra (Ramachandra, Bharath e Sowmyashree, 2014) apresenta no seu trabalho uma forma diferente de compreender as alterações existentes no espaço urbano, através da utilização de imagens de satélite¹⁵. Com base na análise destas imagens, neste caso em particular observando as

¹⁵ Neste trabalho foram usados dados provenientes da rede de satélites Landsat

transformações nas zonas rurais ao nível da ocupação do espaço, conseguem monitorizar de que forma as cidades estão ou não a crescer e para que zonas as mesmas tendem a crescer. Anteriormente Taubenbock (Taubenböck *et al.*, 2012) tinha desenvolvido idêntico trabalho para diferentes cidades utilizando a mesma rede de satélites complementada por mais satélites da rede alemã TerraSAR-X.

Outra tipologia de infraestruturas criadas especificamente para a observação dos movimentos existentes dentro de um espaço urbano são os **loops indutivos**, instalados nas estradas que se pretendem monitorizar. Tirando partido da interpretação das variações no campo magnético produzido por estes *loops* e influenciado pela estrutura metálica dos automóveis, é possível identificar quando um veículo se encontra sobre este sensor. Com a identificação da presença/passagem de um veículo, este sensor permite assim contabilizar o número de veículos que realizaram passagens em determinada via de comunicação (Perttunen *et al.*, 2015). A Figura 2.26 ilustra de que forma a passagem de três veículos sobre um destes sensores altera significativamente o campo magnético, permitindo com base na variação deste campo detetar quando ocorre a passagem de um veículo.

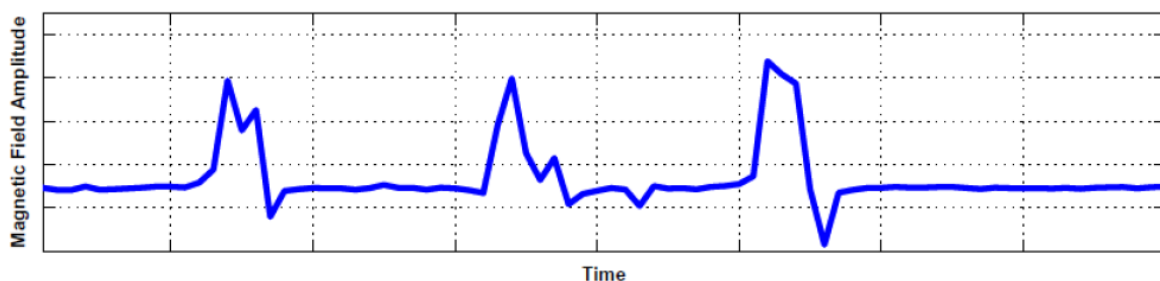


Figura 2.26 – Assinatura magnética após três passagens de veículos (Ahdi et al., 2012)

Bacon (Bacon *et al.*, 2011) refere além dos *loops* indutivos uma outra tecnologia usada como infraestrutura fixa para a realização de contagem de veículos em vias de comunicação, os **detetores de infravermelhos**. Estes detetores quando instalados numa via de comunicação permitem a identificação de veículos, permitindo assim saber o número de veículos que circulou num determinado período temporal em cada uma das direções.

A Figura 2.27 apresenta do lado esquerdo o resultado da deteção automática dos veículos com recurso a infravermelhos e no lado direito o comparativo real sobre o estado da via de comunicação no momento. Como Bacon refere ainda no trabalho, como os detetores de infravermelhos ofuscam os detalhes mais importantes dos veículos (como os seus ocupantes e a matrícula), este método de observação de mobilidade garante a privacidade dos indivíduos observados.



Figura 2.27 – Detecção automática de veículos por infravermelhos (Bacon et al., 2011)

Uma outra forma de realizar a contagem de veículos num espaço urbano passa pela instalação de **câmaras de vigilância** com mecanismos de identificação e contagem de veículos (Figura 2.28) (Koller, Daniilidis e Nagel, 1993). A instalação destas câmaras nas várias vias de comunicação permite não só monitorizar e facilitar a gestão de tráfego, como produzir dados interessantes que podem ser usados para a compreensão da mobilidade dos veículos ao longo do espaço urbano.



Figura 2.28 – Exemplo da identificação de veículos com recurso a imagem vídeo (Koller, Daniilidis e Nagel, 1993)

Dodge (Dodge e Kitchin, 2006) refere outro cenário de utilização de imagens de vídeo obtidas a partir das redes de câmaras instaladas em diferentes cidades britânicas. Neste caso são usados sistemas automáticos de reconhecimento de matrículas com o objetivo de reduzir o congestionamento de tráfego no centro da cidade de Londres através da implantação de um sistema de pagamento. Neste

caso em particular estamos perante um misto de portagens e câmaras de vigilância. Tanto no exemplo de Koller como de Dodge, a utilização das câmaras de vigilância para a realização de contagem de veículos é muito semelhante, o que os distingue é a metodologia de identificação dos veículos. No primeiro caso é identificada a forma do veículo enquanto no segundo caso o reconhecimento passa pela forma da matrícula e seu conteúdo, uma vez que é relevante para aquele tipo de sistema a associação da matrícula ao proprietário do veículo.

Embora tenhamos até agora visto um conjunto alargado de tipologias de dados com potencial para o seu uso na observação do movimento humano, é normal que pelas características das análises que se pretendem realizar ou pela maior ou menor dificuldade em obter os dados, ou mesmo a inexistência de determinada informação, seja necessário o desenvolvimento de aplicações (onde as pessoas são envolvidas no processo de recolha de observações do seu próprio movimento) ou sistemas destinados principalmente à recolha de informação de mobilidade.

Com o objetivo de organizarmos adequadamente os conceitos relativos à recolha de dados baseada em aplicações é importante fazer nesta fase a distinção entre dois conceitos de observação/sensorização: participativo e oportunista. Lane (Lane e Eisenman, 2008) apresenta de forma muito clara o que distingue estes dois conceitos: uma colaboração participativa implica que o utilizador de forma consciente opte por utilizar uma aplicação ou dispositivo, sabendo de antemão que os seus dados vão ser partilhados e que o utilizador poderá ter algum retorno (pessoal ou financeiro) nessa partilha; pelo contrário, numa colaboração oportunista o utilizador pode nem ter conhecimento que os seus dados são partilhados, uma vez que o processo de recolha de dados não exige por parte do utilizador nenhuma intervenção manual, e muitas vezes nem sequer tem conhecimento que a aplicação que o está a observar está no momento ativa.

Existem inclusivamente processos de recolha oportunista de dados em que o utilizador não está consciente sobre a recolha e partilha de dados, porque a informação sobre o processo está “escondida” no meio dos longos textos que constituem as condições de utilização de certas aplicações ou sistemas operativos para dispositivos móveis.

Ao nível da interação exigida ao utilizador também existem diferenças: no primeiro caso existe uma interação explícita com vista à recolha de dados, enquanto no segundo caso não é necessária uma intervenção explícita do utilizador, até porque a existir necessidade de interação, o utilizador seria alertado para a execução de uma determinada aplicação sem o seu conhecimento.

Como refere e bem Lane , este último tipo de colaboração deve ter em atenção duas questões importantes: a privacidade dos utilizadores, uma vez que estão a ser recolhidos e partilhados dados que podem ser considerados pelos utilizadores como informação sensível (até porque a recolha dos dados é feita de forma dissimulada e como tal o utilizador pode não ter conhecimento sobre que dados estão a ser recolhidos) e o impacto que este tipo de sistemas de observação têm nas rotinas do utilizador (não é conveniente que afete ou altere o dia-a-dia do utilizador).

Um dos exemplos de aplicações desenvolvidas especificamente para a recolha de dados, orientada para a análise de mobilidade humana é a *MOVE*¹⁶. Esta aplicação desenvolvida pela Universidade de Coimbra para dispositivos *Android* tem como principal objetivo a recolha de três tipos de dados: GPS, Wi-Fi e rede celular. Como esta aplicação é instalada em smartphones, e como estes acompanham os indivíduos nas suas atividades diárias, os dados recolhidos são uma representação fiel das movimentações realizadas pelos indivíduos. O grande inconveniente desta aplicação, assim como de outras para o mesmo efeito, é o consumo energético. A utilização periódica destes três sensores obriga a que os mesmos estejam permanentemente ligados, diminuindo assim a autonomia dos equipamentos. Esta aplicação, como apenas recolhe dados com um fim específico e implica o conhecimento do utilizador que a instala e utiliza, poder-se-ia enquadrar assim na observação participativa, no entanto como não obriga a uma intervenção explícita dos utilizadores também pode ser considerada de oportunista. Dentro deste conjunto de aplicações importa referir como outros exemplos as aplicações: *LifeMap* (Chon e Cha, 2011), *SenseMyCity*¹⁷, *Radwende*¹⁸ e *WeCycle*¹⁹. Enquanto as duas primeiras aplicações recolhem informação relativamente ao contexto envolvente do utilizador, as duas últimas são responsáveis por recolher dados de mobilidade apenas direcionados aos ciclistas de forma a permitir determinar padrões de mobilidade e nesse sentido garantir que as alterações ou adaptações futuras que existam no espaço urbano tenham em conta as necessidades desta alternativa ecológica de mobilidade.

¹⁶ Aplicação para dispositivos Android gentilmente cedida pela Universidade de Coimbra com vista à recolha de dados de mobilidade.

¹⁷ Disponível em: <http://cloud.futurecities.up.pt/sensemymcity/> (visitado em Agosto de 2016)

¹⁸ Disponível em: <https://www.radwende.de/en/> (visitado em Agosto de 2016)

¹⁹ Disponível em: <http://www.travelai.info/wecycle-safer-cycling.html> (visitado em Agosto de 2016)

Em contraste com as aplicações de observação participativa está por exemplo a aplicação *GeoAnuncios*²⁰ que de forma oportunista recolhe dados espaçotemporais. Esta aplicação desenvolvida na Universidade do Minho e cuja principal funcionalidade para o utilizador é a publicação e pesquisa de anúncios com base na georreferenciação tanto dos anúncios como do utilizador, guarda também dados de localização temporal e espacial dos utilizadores.

Mas não são apenas as aplicações desenvolvidas em ambiente académico e com fins meramente investigacionais que de forma dissimulada recolhem dados, neste caso dados com aplicabilidade nos desafios atuais de compreensão e padronização da mobilidade humana.

A Google, por exemplo, monitoriza a nossa localização recolhendo dados nas várias aplicações e serviços que disponibilizam com claros objetivos comerciais, em particular, para a apresentação contextualizada de anúncios. Não se pode dizer que esta recolha seja participativa, antes pelo contrário, é oportunista e dissimulada uma vez que a informação ao utilizador sobre a recolha de dados está também ela “incompleta” nas configurações da nossa conta. A informação sobre o histórico de localizações apresentada pela Google (Google, 2015) refere:

“O Histórico de localizações cria um mapa privado dos locais para onde se desloca com os seus dispositivos com sessão iniciada para obter melhores pesquisas no mapa, trajetos para deslocamentos diários e muito mais.”

No entanto a mais-valia deste histórico não é de todo apenas para o utilizador. De qualquer forma podemos sempre cancelar a partilha das nossas localizações com o Google e ainda consultar a informação recolhida no *Google Location History* (Google, 2015).

Já vimos que a recolha de dados por aplicações pode e é feita com maior ou menor conhecimento do utilizador. Mas possivelmente o que a maioria dos utilizadores de dispositivos móveis não sabe é que os próprios **sistemas operativos móveis** monitorizam a nossa localização.

Mais uma vez com o argumento de mais-valia para o utilizador, neste caso para o *iOS* (Apple Inc., 2015): *“Por exemplo, uma aplicação poderá utilizar os seus dados de localização e as consultas de localização para o ajudar a encontrar cafés ou cinemas nas proximidades. O dispositivo poderá ainda definir automaticamente o fuso horário com base na sua localização atual.”*

²⁰ Disponível em: <http://geoanuncios.dsi.uminho.pt/> (visitado em Agosto de 2016)

Certo é que neste caso a Apple também tem interesse nas nossas localizações (Apple Inc., 2015): *“Se os Serviços de localização estiverem ativos, o dispositivo irá enviar periodicamente as localizações geográficas de hotspots Wi-Fi e torres de comunicações nas imediações para a Apple, de uma forma anónima e cifrada, para aumentar a base de dados de hotspots Wi-Fi e torres de comunicações. Para além disso, se estiver em viagem (por exemplo, num automóvel) e os Serviços de localização estiverem ativos, um dispositivo iOS com GPS ativado irá também enviar periodicamente localizações de GPS e informações relativas à velocidade para a Apple, de uma forma anónima e cifrada, para aumentar a base de dados de trânsito. Os dados de localização recolhidos pela Apple não identificam pessoalmente o utilizador.”*

Como seria de esperar esta prática não é exclusiva do *iOS*, no caso do sistema operativo *Android* o procedimento de recolha de dados mantém-se, estando neste caso mais acessível o acesso aos dados diretamente a partir do equipamento. Neste sistema operativo o acesso ao sistema de ficheiros é facilitado comparativamente ao *iOS*, onde o acesso não é permitido. No caso do sistema operativo *Android* os dados de localização encontram-se em:

/data/data/com.google.android.location/files

onde estão disponíveis os ficheiros com os dados contendo informação sobre as redes Wi-Fi, por exemplo:

key	Accuracy	conf.	latitude	longitude	time
50:63:13:57:42:7e	80	92	57.689354	11.994763	04/11/11 10:03:51
e0:cb:4e:7e:cc:53	75	92	57.689340	11.994495	04/11/11 10:03:51

Tabela 2.8 – Exemplo dos dados Wi-Fi recolhidos²¹

ou informação sobre a rede celular:

key	accuracy	conf.	latitude	Longitude	time
240:5:15:983885	1186	75	57.704031	11.910801	04/11/11 20:03:14
240:5:15:983882	883	75	57.706322	11.911692	04/13/11 01:41:29

Tabela 2.9 – Exemplo de dados da rede celular²²

²¹ Dados extraídos de: <https://github.com/packetss/android-locdump> (visitado em Agosto de 2016)

²² Dados extraídos de: <https://github.com/packetss/android-locdump> (visitado em Agosto de 2016)

A Figura 2.29 ilustra um resumo dos vários exemplos de formas de observação do movimento humano.

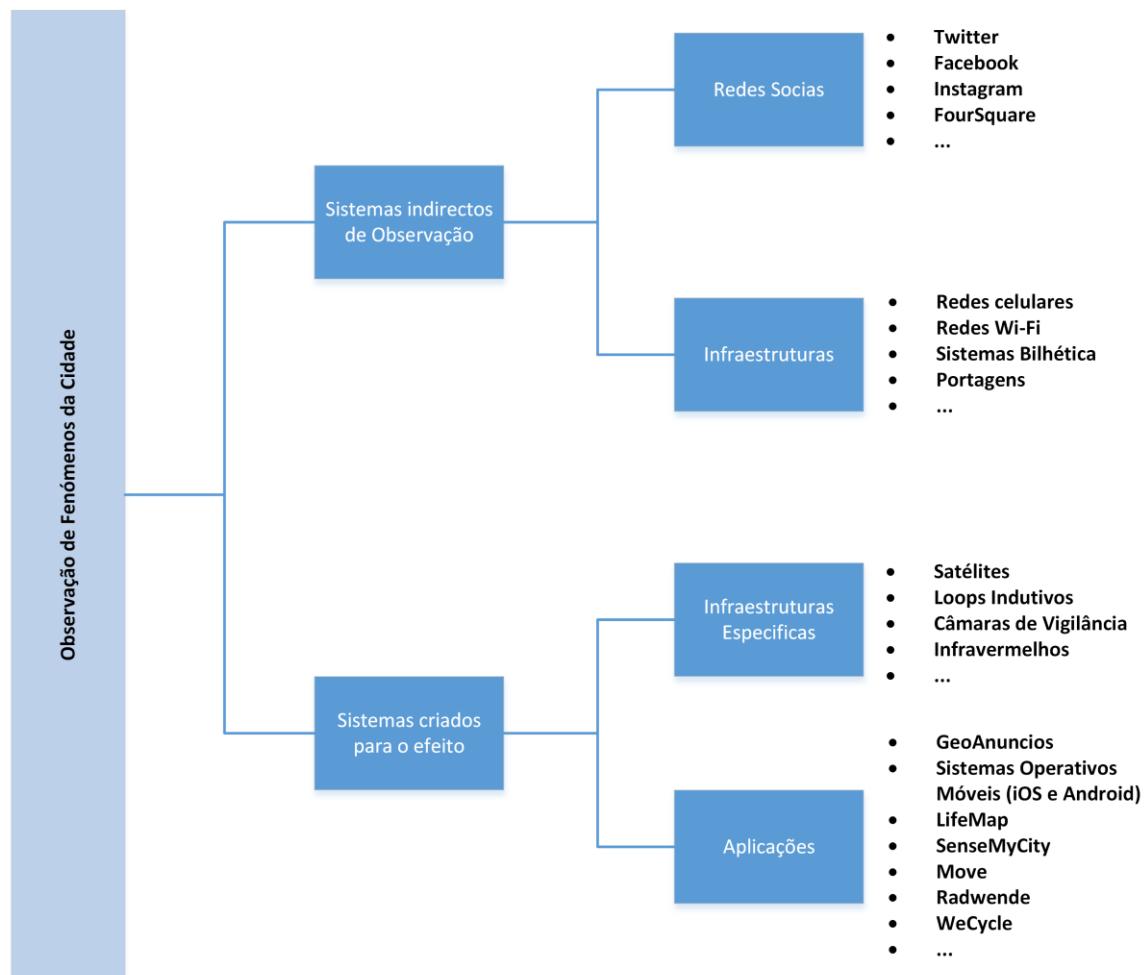


Figura 2.29 – Tipologias de observações

Existirão com toda a certeza outros exemplos quer para os tipos apresentados ao longo desta seção, assim como para outras formas de observarmos o movimento humano (e.g. a circulação de notas bancárias). Esta diversidade de observações origina uma diversidade nos dados recolhidos, existindo por isso dados e respetivos atributos, com características diferentes dependentes das tecnologias utilizadas na observação dos artefactos.

2.5. Técnicas de análise

Planear um espaço urbano requer um mínimo de conhecimento sobre as características da mobilidade humana. Compreender se existe algum relacionamento entre locais, se a maioria dos

movimentos é de curta ou longa distância, quais os momentos temporais em que ocorrem a maioria desses movimentos, entre outras características, são conhecimentos fundamentais para uma adequada atuação.

Uma das formas de obtermos então este conhecimento é através de **análise estatística** dos dados de mobilidade disponíveis. Foi exatamente isso que Brockmann (Brockmann, Hufnagel e Geisel, 2006) fez ao analisar estatisticamente a mobilidade humana com base neste caso particular na movimentação de notas bancárias ao longo dos Estados Unidos, permitindo assim extrair conclusões relativamente à natureza do movimento humano em espaços geográficos alargados. Dois anos depois González (González, Hidalgo e Barabási, 2008) recorrendo também à análise estatística, neste caso de dados de redes celulares, conseguiu extrair conclusões interessantes nos padrões de mobilidade humana para pequenas e longas distâncias. Além de análises de padrões para áreas geográficas de dimensão nacional ou citadina, Lind (Lind e Moreira, 2015) realizou um estudo baseado em análise estatística de onde extraíram padrões para distâncias muito curtas, principalmente entre edifícios de *campus* universitários.

Os padrões de mobilidade humana identificados a partir de análise estatística permitem obter uma visão global dos comportamentos dentro das cidades e entre cidades. No entanto, possivelmente não são das técnicas de análise mais adequadas se pretendermos obter, por exemplo, os locais relevantes, individuais (Zhou *et al.*, 2007) ou globais (Andrienko *et al.*, 2011).

Marmasse (Marmasse e Schmandt, 2000) realizou possivelmente um dos primeiros trabalhos que explorava a descoberta de locais relevantes (*places*) realizando a deteção dos mesmos com base na perda de sinal GPS. Se o sinal GPS estivesse ausente durante três períodos temporais consecutivos, o sistema marcava a próxima posição GPS válida como a localização a considerar. Como os autores constataram, a perda de sinal GPS era frequente em zonas urbanas, onde os edifícios dificultam a visibilidade dos satélites, dificultando assim a aquisição de coordenadas válidas de GPS e por consequência a correta inferência de locais. A extração de locais neste trabalho não contemplava grandes mecanismos de agregação, contrariamente ao trabalho realizado anos depois por Ashbrook (Ashbrook e Starner, 2002) que, também com recurso a dados de GPS, recolhidos durante quatro meses para um único utilizador, desenvolveu um algoritmo de **clustering**, uma variante do algoritmo *k-means*, que permitiu extrair os locais relevantes para esse mesmo utilizador com base em dois critérios de distância: raio e tempo. Uma nova coordenada GPS só seria adicionada a um *cluster* se a

distância geográfica fosse inferior ao raio estipulado e se a distância temporal fosse inferior a 10 minutos; caso contrário, seria criado um novo *cluster*.

Usando também a extração de *clusters* como técnica de análise, Kang (Kang *et al.*, 2005) concluiu ser importante que a agregação dos acontecimentos fosse baseada não só na componente geográfica, mas também no tempo, isto porque verificaram que existem dois tipos de locais relevantes para os indivíduos: aqueles onde passamos a maior parte do nosso tempo e os que visitamos frequentemente com reduzidos tempos de permanência. Por este motivo os autores consideraram ser necessário uma distinção no momento de agregação para ser possível contemplar estas duas situações, devendo assim serem usados valores distintos para o critério temporal.

Kang (Kang *et al.*, 2005) concluiu ainda que os algoritmos tradicionais de *clustering* não são os mais adequados a este tipo de análises, com componentes temporais. É assim comum que sejam feitas adaptações aos algoritmos originais ou que a sua utilização seja feita integrando estes algoritmos de *clustering* em outros algoritmos. Foi o que Andrienko (Andrienko *et al.*, 2011) fez quando elaborou um algoritmo para cálculo da distância espaçotemporal e aplicou o algoritmo de *clustering* a essa função de distância. Neste último trabalho os dados importantes para a agregação são o tempo e o espaço. Num outro trabalho, Andrienko (G. Andrienko, Andrienko, Bosch, *et al.*, 2013) recorreu novamente a algoritmos de *clustering* para extrair locais relevantes não pelos acontecimentos em si, mas pela agregação de palavras extraídas dos *tweets*. Com base na identificação de termos provenientes dos *tweets* e respetiva agregação, foi possível aos autores caracterizar ao nível semântico o espaço, identificando zonas de serviços, habitacionais, entre outros. Estamos neste caso perante um objetivo diferente da extração de locais relevantes, a caracterização automática do espaço. Santos (Santos *et al.*, 2012) no âmbito do projeto TICE.Mobilidade realizou semelhante trabalho, onde procurou enriquecer semanticamente locais considerados relevantes (Pontos de Interesse) com base em informação disponível na Web.

Estes dois últimos trabalhos de Andrienko, além de contemplarem análises baseadas em *clustering*, consideram ainda, e possivelmente é o aspeto mais relevante nestes dois trabalhos, uma técnica de análise baseada na visualização e manipulação dos dados. Esta técnica de análise denominada **visual analytics** tem como objetivo permitir de forma manual a manipulação dos dados disponíveis de forma a realizarmos um conjunto de validações ou testes de hipóteses. Esta manipulação dos dados, quando realizada por um especialista, permite a realização de um conjunto de hipóteses confirmadas ou não através da inspeção visual. É natural que por vezes não exista uma

clara definição das hipóteses a testar que, no entanto, com a validação visual e consequente modificação dos parâmetros permite a descoberta ou confirmação de hipóteses. Como refere Thomas (Thomas e Cook, 2005), esta técnica de análise recorre a ferramentas para sintetizar informações em análises visuais procurando identificar o esperado e descobrir o inesperado.

No trabalho de MacEachren (MacEachren e Robinson, 2011) é apresentada uma ferramenta de *visual analytics* apelidada *SensePlace*²³ que permite de forma visual explorar hipóteses variando, neste caso em particular, os termos de pesquisa associados às publicações na rede social *Twitter*. A construção dos elementos a pesquisar poderá ter em conta a identificação de um local e/ou uma data. Ferreira (Ferreira, Poco e Vo, 2013) apresenta no seu trabalho o *TaxiVis*, uma ferramenta de *visual analytics* que utilizando dados das viagens de táxis permite a manipulação dos mesmos no sentido de ser possível a exploração visual de diferentes combinações e/ou hipóteses relativas à atividade dos táxis. Os autores acreditam que os padrões de mobilidade dos táxis refletem em certa medida os padrões globais das cidades.

De acordo com Kim (Kim *et al.*, 2009) os algoritmos de ***place learning*** têm como objetivo descobrir locais importantes para um indivíduo ao qual está associado um significado semântico, por exemplo: “casa”, “trabalho”, entre outros. Anteriormente, quando referimos os algoritmos de *clustering*, tínhamos já visto trabalhos onde o objetivo era idêntico, a extração mais ou menos automática de locais relevantes para os indivíduos. No entanto existe uma grande diferença entre os algoritmos de *clustering* e os algoritmos de *place learning*. Estes últimos têm como *input* os dados obtidos por diversos sensores de acordo com uma ordem cronológica e como *output* uma estrutura de dados que poderá ser composta, por exemplo, pelos momentos temporais de entrada e saída desse local (Kim *et al.*, 2009). Assim, um “place” não é apenas caracterizado por um conjunto de atributos espaciais, mas também por atributos temporais (duração das estadias, tempo médio de estadia, perfil temporal de ocupação, etc.). Ao contrario dos exemplos de algoritmos de *clustering* cuja agregação era

²³ Disponível em: <http://www.geovista.psu.edu/SensePlace2/> (visitado em Agosto de 2016)

baseada em distâncias temporais e/ou geométricas, nos casos dos algoritmos de *place learning* a agregação/associação é feita de acordo com a semelhança a uma *fingerprint*²⁴.

No caso do trabalho de Kim (Kim *et al.*, 2009) é apresentado o *PlaceSense*, um algoritmo de *place learning* que utiliza informação Wi-Fi ou celular para inferir os momentos de entrada e de saída de um local. No caso dos dados Wi-Fi, as *fingerprints* recolhidas têm na sua composição os diferentes endereços MAC dos vários APs visíveis.

Embora exista já trabalho significativo na área de análise do movimento humano, verifica-se um leque muito restrito de abordagens, donde as que recebem maior destaque são as quatro apresentadas anteriormente.

2.6. Desafios ainda sem resposta

Ao longo deste capítulo verificamos a existência de diferentes formas de observação, a que correspondem diferentes tipos de dados, com níveis variáveis de resolução espacial e temporal, assim como diferentes formas de visualização e análises de padrões de mobilidade humana. No entanto, existem desafios e questões que necessitam ainda de atenção e que por isso merecem ser explorados ao nível académico.

A quantidade de dados recolhidos para uma análise de mobilidade humana, mesmo que seja de um espaço urbano com uma dimensão reduzida, resulta em *datasets* de elevadas dimensões para que os mesmos possam ser usados ou trabalhados em tempo real de forma ágil e podendo até originar questões quanto ao seu registo físico em disco ou outros dispositivos. Como MacEachren (MacEachren e Robinson, 2011) refere, os dados produzidos pelas redes sociais, como exemplo, são em tão grande número que o desafio passa por conseguir realizar uma análise a um volume tão elevado de dados. Temos assim um desafio em aberto, o arquivo e processamento de **grandes volumes de dados**.

²⁴ *Fingerprint* é neste contexto uma amostra composta por dados de um ou mais sensores, onde a estrutura de dados de cada sensor é regra geral muito própria e com reduzidas semelhanças entres sensores, e que caracterizam de forma única um determinado local.

Mas esta problemática só acontece quando efetivamente temos dados, porque a ausência ou inadequação dos dados também são em si um problema. Caso não existam, poderemos sempre desenvolver aplicações ou sistemas para os recolher, com o esforço de desenvolvimento e limitações de escala associadas (poderemos ter disponível uma amostra reduzida não sendo por isso conclusiva a análise). A inadequação também é um problema quer por faltarem características aos dados ou por os dados não estarem atualizados. A utilização de questionários ou dados obtidos pelos Censos, por exemplo, podem ter uma antiguidade tal que a análise que se pretende realizar poderá não fazer sentido. Portanto um desafio ainda existente é o **acesso ou disponibilidade** de dados em quantidade e qualidade para análise de mobilidade humana, quer no momento da recolha dos dados ou mesmo no pré-processamento dos dados. Muitas vezes os dados existentes resultam então dos questionários e, pelo facto de estes estarem em papel e não em formato eletrónico, exige logo à partida um esforço extra. Jiang (Jiang, Ferreira e Gonzalez, 2012) sentiu esta dificuldade quando realizou um estudo de mobilidade, uma vez que usaram informação recolhida de forma manual proveniente de diários das viagens realizadas pelos habitantes de Illinois.

A problemática acessibilidade aos dados poderia em parte ser resolvida com a massificação das iniciativas *open data*, como acontece por exemplo em cidades como Nova Iorque²⁵ ou Londres²⁶. No entanto, pelo menos em Portugal onde estas iniciativas não são ainda uma realidade, torna-se demasiado burocrático senão mesmo impossível conseguir o acesso a informação existente e que faria todo o sentido ser pública. Um exemplo flagrante desta realidade aconteceu ao longo do projeto mobilizador TICE.Mobilidade²⁷ onde os maiores operadores na área de transportes mesmo tendo dados relevantes para o projeto, e com claros benefícios para os próprios, não mostraram abertura para a sua disponibilização.

A recolha de dados com base em diferentes formas de observação da mobilidade humana, assim como as diferentes técnicas de análise e de visualização, dificultam a comparação de resultados obtidos por diferentes investigadores. A criação de uma **estrutura uniforme** para a representação

²⁵ NYC OpenData: <https://nycopendata.socrata.com/> (visitado em Agosto de 2016)

²⁶ London Datastore: <http://data.london.gov.uk/> (visitado em Agosto de 2016)

²⁷ Mais informação em: <http://tice.mobilidade.ipn.pt/> (visitado em Agosto de 2016)

da informação de mobilidade com suporte para dados espaçotemporais obtidos por múltiplos e distintos sensores torna-se assim um desafio, relevante tanto para a extração de conclusões sobre a mobilidade, como para o suporte à visualização coerente da informação.

Capítulo 3. Ecosistema Urbano

No capítulo anterior estão descritos alguns exemplos de observações de mobilidade humana recorrendo a diferentes sensores e consequentemente diferentes tipologias de dados (Figura 3.1). Com maior ou menor dificuldade na extração e uso desses dados, verificamos ser possível inferir algumas conclusões acerca dos padrões de mobilidade e identificação de potenciais dificuldades ou constrangimentos existentes no espaço urbano.



Figura 3.1 - Observação e sensores

3.1. Observação e Análise

A observação em si de artefactos²⁸ pode, de forma superficial, resumir-se apenas à recolha de dados provenientes de um ou mais sensores, como por exemplo: GPS, rede celulares, redes sociais, Wi-Fi, entre outros. A observação consiste assim em identificar através dos dados recolhidos onde e quando um determinado artefacto foi detetado. Os dados utilizados para a observação de um artefacto, podem não estar, e muitas vezes não estão, num formato adequado para servirem de base ao estudo da mobilidade humana, sendo por isso necessário uma fase de processamento e transformação desses mesmos dados (nos capítulos seguintes esta fase é alvo de maior detalhe tendo em conta o trabalho desenvolvido). Se a informação incluída nos dados obtidos pela observação é relevante para a análise que se pretende realizar, são igualmente relevantes as fases de tratamento e processamento dos dados, uma vez que culminarão na criação de mapas/vistas de mobilidade individual ou global.

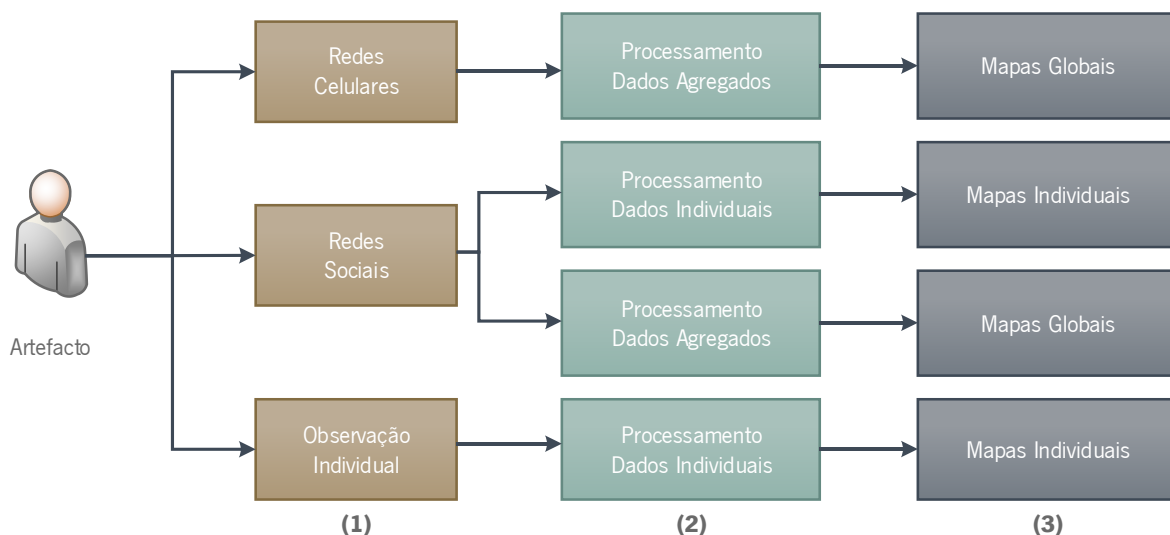


Figura 3.2 – Exemplos do processo de criação de mapas de mobilidade

Embora exista à data uma quantidade relevante de trabalhos direccionados para a análise de mobilidade humana, os mesmos ficam um pouco aquém do que seria a visão completa da mobilidade humana existente no espaço urbano, isto é, alguns dos trabalhos analisam a mobilidade de forma

²⁸ Denominamos de artefactos os objetos e/ou indivíduos observados. Muitas vezes não observamos em si os indivíduos, mas outros objetos, como por exemplo: autocarros, táxis, bilhetes, publicações em redes sociais, entre outros.

parcial não contemplando ou representando toda a verdadeira mobilidade existente nesse mesmo espaço.

A Figura 3.2 ilustra, com base nos trabalhos apresentados no capítulo anterior, alguns exemplos das várias fases do processo de criação de mapas de mobilidade. Tendo em consideração os dados recolhidos a partir da observação dos artefactos (1) e de acordo com as próprias características dos mesmos, sejam eles dados individuais ou dados agregados, o seu processamento (2) e consequente extração de mapas individuais ou globais (3), produz apenas uma visão parcial dos diferentes fenómenos de mobilidade existentes no espaço urbano. Aprofundando apenas como exemplo o trabalho realizado por Reades (Reades *et al.*, 2007), as observações obtidas neste caso têm origem em dados agregados da utilização de redes celulares não permitindo uma distinção clara sobre que utilizadores da rede foram ou estão a ser observados, sendo impossível a produção de mapas de mobilidade individuais a partir deste sensor, pelo que apenas será possível a criação de mapas de mobilidade globais.

Da mesma forma que as características dos dados influenciam o tipo de mapas produzidos, também a tipologia dos mapas gerados influencia a análise e interpretação da mobilidade humana. A Figura 3.3 apresenta, de forma não exaustiva, como os dados e mapas influenciam o estudo que se pode realizar.

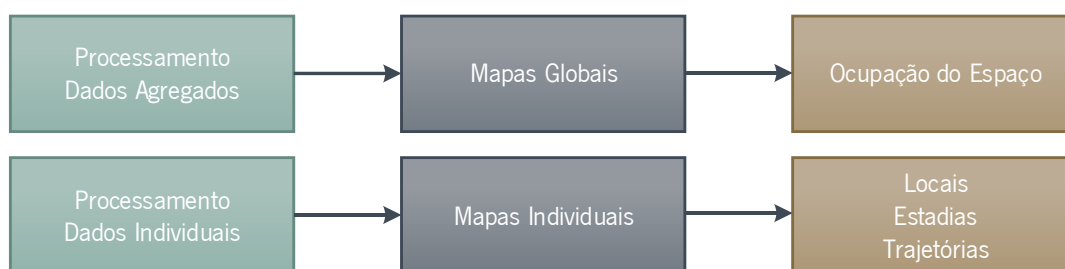


Figura 3.3 – Representação dos resultados da análise de mobilidade

Como os mapas individuais e globais têm informação distinta, produzem consequentemente resultados distintos e conclusões diferentes. Enquanto os mapas globais obtidos a partir de dados agregados permitem ilustrar de que forma os artefactos estão distribuídos ao longo do espaço urbano (representando zonas de maior ou menor densidade populacional), já os mapas individuais potenciam a extração de informação mais detalhada, como por exemplo: os locais visitados pelo indivíduo; as estadias que realizou nesses locais; ou as trajetórias realizadas entre locais (com maior ou menor rigor geométrico dependendo do(s) sensor(es) utilizado(s)).

Os exemplos dos resultados obtidos de acordo com a tipologia dos mapas, como ilustra a Figura 3.3, podem facilmente ser verificados com dois dos trabalhos apresentados no capítulo anterior. Mais uma vez recorrendo ao trabalho de Reades (Reades *et al.*, 2007) verificamos um excelente exemplo de ilustração da ocupação do espaço urbano com base nas observações; enquanto por exemplo o trabalho de Andrienko (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013) que recorre à representação através de cubos espaço-tempo para a produção de mapas individuais de mobilidade, permite a extração de locais visitados pelos indivíduos, as suas estadias, bem como as trajetórias realizadas entre os diferentes locais visitados. Temos assim diferentes resultados em função da tipologia dos mapas de mobilidade gerados, mapas estes dependentes das características das observações recolhidas.

Estudar a mobilidade humana ao longo do espaço urbano partindo apenas da ocupação do espaço tem as suas vantagens, como por exemplo: perceber locais de maior densidade de indivíduos com ou sem componente temporal. No entanto dificulta a perceção de movimentos no espaço, uma vez que não existe nesta abordagem uma descrição da origem e destino dos movimentos dos indivíduos. Ambas as componentes são relevantes para a compreensão quer da mobilidade humana, quer dos seus reflexos na qualidade de vida dos cidadãos, não sendo por isso aconselhável a sua dissociação ou a sua utilização apenas parcial.

Só com o recurso a mapas individuais de mobilidade é possível identificar os locais relevantes para o indivíduo e as trajetórias realizadas entre esses locais. A utilização de apenas mapas globais obtidos a partir unicamente de dados agregados ofuscam fatores e características da mobilidade humana. No entanto os mapas individuais analisados de forma isolada (se analisados um a um) não demonstram a complexidade da mobilidade em termos da densidade dos efeitos da mobilidade dos indivíduos num referencial espaçotemporal. Não é a movimentação de um ou dois indivíduos que geram dificuldades, mas sim quando um conjunto alargado de indivíduos, num mesmo referencial espaçotemporal, realizam semelhantes movimentações.

Com base nas vantagens e desvantagens que ambos os mapas apresentam, a abordagem com maior potencial para agregar no mesmo processo as mais-valias dos mapas individuais e mapas globais deveria incluir as seguintes fases:

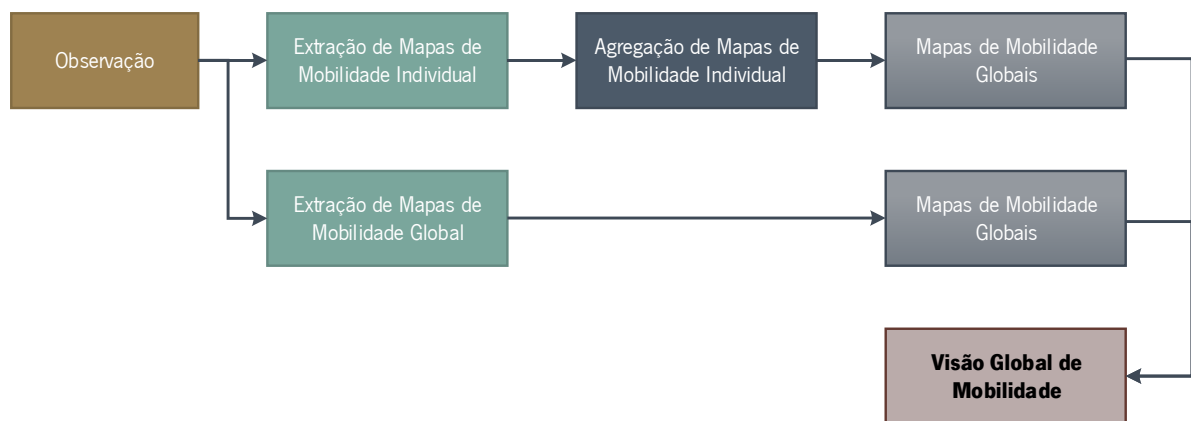


Figura 3.4 - Proposta para a elaboração de estudos de mobilidade

A proposta apresentada na Figura 3.4 visa clarificar os passos/processos mínimos necessários para a criação de estudos de mobilidade que permitam a obtenção de uma visão global, o mais completa possível, do espaço urbano. Esta visão deverá contemplar sem dúvida os mapas de mobilidade globais obtidos a partir de dados agregados, mas também contemplar mapas de mobilidade globais obtidos pela agregação de mapas individuais de mobilidade. A inclusão no mesmo estudo de mapas obtidos através de dados agregados e de observações individuais de artefactos permite enriquecer a informação que está em análise potenciando a extração de conclusões mais ricas, mais próximas da realidade e mais abrangentes. É através da conjugação dos pontos fortes destes dois tipos de mapas na mesma análise que se consegue ter uma descrição o mais realista possível dos fenómenos associados à mobilidade existentes num espaço urbano.

A observação individual e consequente processamento de mapas individuais de mobilidade têm como seria de esperar uma desvantagem de peso, ser um processo longo e moroso tanto na recolha como no processamento para conjuntos alargados de artefactos suficientes para ser possível a extrapolação de conclusões relevantes. Uma centena de artefactos observados pode não ser relevante para uma cidade de centenas de milhares de habitantes, no entanto poderá ser ilustrativo de alguns padrões existentes, padrões esses que não se revelariam se recorrêssemos apenas a mapas globais baseados em dados agregados. Por exemplo, imaginando que temos mapas individuais de mobilidade de um conjunto de alunos e de um conjunto de docentes da Universidade do Minho, muito possivelmente conseguimos identificar padrões comuns na mobilidade de e para a Universidade, mas distintos nos restantes momentos temporais. Se a este exercício adicionarmos um mapa individual de um outro indivíduo que não interaja com a Universidade do Minho, com certeza que identificaremos um perfil completamente distinto dos dois anteriores. A identificação de diferentes perfis de mobilidade

é apenas possível com mapas individuais uma vez que apenas com estes mapas se torna visível a mudança de locais visitados ao longo do espaço.

A comparação de padrões nos mapas individuais não tem que ser apenas entre indivíduos, um mapa individual de mobilidade quando realizado para um período de tempo considerável, permite identificar padrões distintos para o mesmo indivíduo. O caso mais óbvio é o facto de os nossos padrões de mobilidade nos dias úteis não serem equivalentes aos padrões realizados durante os fins de semana. A Figura 2.15 (G. Andrienko, Andrienko, Fuchs, *et al.*, 2013) referenciada na secção 2.3 (Visualização do movimento humano) é a melhor ilustração desta distinção comportamental bem como da própria noção de mapa individual de mobilidade.

Outro facto relevante que justifica o fluxo de processos proposto na Figura 3.4 prende-se com a confiança ou falta dela relativamente aos resultados obtidos a partir da análise de mobilidade tendo por base apenas em observações recolhidas pela utilização, por exemplo, de redes celulares e/ou redes sociais. Em ambas, as observações ocorrem e estão condicionadas pela realização de atividades, sociais ou não, podendo as observações serem esparsas no tempo e no espaço dependendo da frequência com que se realizam as atividades.

Como indica Hasan (Hasan *et al.*, 2013) as grandes fontes de dados relativas a observações de artefactos (redes celulares, uso de cartão de crédito, entre outros) embora sejam uma oportunidade única para identificar padrões de mobilidade humana para conjuntos alargados de indivíduos têm informação muito incompleta sobre as viagens individuais, uma vez que a observação apenas é registada quando se realiza uma atividade, não sendo recolhida a informação de forma contínua.

Em conclusão, para que uma análise de mobilidade relativa a um espaço urbano seja a mais completa possível e reflita ao máximo os comportamentos verdadeiramente existentes, é necessário ter especial cuidado com dois requisitos: os dados recolhidos na observação dos artefactos corresponderem o mais possível à mobilidade diária, isto é, refletirem ao máximo os movimentos e paragens dos indivíduos; e recorrer sempre que possível à dupla abordagem dos mapas globais apresentados na Figura 3.4.

3.2. Atuação

As análises de mobilidade humana feitas quer sobre os mapas individuais, ou através dos mapas globais/coletivos, permitem revelar fenómenos/padrões anómalos ou indesejados existentes

num espaço urbano. Um dos exemplos mais óbvio são os congestionamentos de tráfego. Ao longo do capítulo anterior surgem visualizações onde os congestionamentos de tráfego estão bem presentes. Assim, ao analisarmos a mobilidade humana conseguimos muito possivelmente extrair conclusões sobre, por exemplo, os períodos do dia e áreas geográficas específicas onde existem dificuldades resultantes do congestionamento de tráfego.

No entanto, as análises de mobilidade embora cumprindo as suas funções, limitam-se a demonstrarem algo como: congestionamento de tráfego nos locais X e com maior impacto nos períodos temporais Y. Identificam e bem as anomalias, mas não as causas das mesmas e como tal dificilmente se conseguem corrigir. Sabemos assim, **quando e onde** o problema ocorre, mas muito dificilmente saberemos o **porquê** do problema ocorrer e o **como** o resolver. A correção ou minimização das anomalias e respetivos efeitos nas vidas dos cidadãos deve ser feita por quem tem o poder de decisão ou pelo menos o poder de influenciar a decisão.

É neste momento que deverá entrar a atuação sobre o espaço urbano. Com base nos resultados da observação e análise da mobilidade humana, o **planeador/atuador/decisor** deverá atuar de forma a explorar soluções que no mínimo minimizem os problemas ou que no melhor dos casos eliminem o problema.

Embora seja uma prática pouco ou nada explorada, a utilização dos resultados da observação como apoio nas tomadas de decisão relativamente ao planeamento atual e futuro, das nossas cidades em particular e do espaço urbano em geral, será sempre uma mais-valia para uma adequada atuação. É sempre preferível que as tomadas de decisão sejam sustentadas por informação, de preferência a mais atual possível, de modo a que a resposta ao nível do planeamento vá ao encontro das atuais e reais necessidades dos cidadãos que frequentam esses espaços urbanos.

Infelizmente para todos nós, são raros os exemplos em que existe uma consciencialização sobre a forma adequada de planear de acordo com as reais necessidades. Um desses exemplos foi implementado em Londres através do desenvolvimento de uma aplicação móvel, *WeCycle*^{29,30}, que

²⁹ <http://www.fastcoexist.com/3034313/a-new-app-tracks-where-cyclists-actually-ride-to-help-plan-better-paths> (visitado em Agosto de 2016)

³⁰ <http://road.cc/content/news/126952-wecycle-app-aims-provide-planners-useful-cycling-data-free> (visitado em Agosto de 2016)

observava os padrões de mobilidade dos ciclistas com o objetivo de colmatar a falta de informação sobre as necessidades deste público-alvo em particular, visando o planeamento futuro de novas ciclovias. Este é um dos bons exemplos em que se procurou atuar no espaço urbano de acordo com as necessidades observadas.

3.3. Ecossistema de Mobilidade Urbana

O planeamento urbano é sem dúvida uma tarefa árdua e complexa. Primeiro, porque envolve normalmente grandes investimentos; segundo, porque ao modificarmos o espaço urbano irão surgir boas ou más consequências; terceiro, porque a mudança deve ser feita com o menor impacto no que atualmente existe; quarto, porque o ser humano é avesso à mudança; quinto, o planeamento exige equipas multidisciplinares e em alguns casos de diferentes organizações; e poderíamos continuar a enumerar argumentos que sustentam a complexidade do planeamento. Na mesma linha de raciocínio, Hamilton (Hamilton *et al.*, 2005) afirma que as tomadas de decisão sobre o planeamento urbano deverão ter em consideração: a estrutura física da cidade, juntamente com fatores económicos, sociais e ambientais.

No entanto, o planeamento das cidades tem que ser feito e de preferência bem feito. Embora repetida anteriormente a ideia, nunca é demais voltar a referir: não se faz um bom planeamento sem uma boa análise das necessidades e dificuldades atuais. Por esse motivo é fundamental a observação da mobilidade humana.

Um bom exemplo demonstrativo da “falha de comunicação” entre quem planeia/gere, e quem usa o espaço urbano, é o aparecimento de iniciativas como por exemplo: “No meu bairro”³¹, “A minha rua”³² e “Mais cidadão!”³³. Nestas plataformas os cidadãos podem reportar problemas existentes no espaço urbano, sendo uma forma de apresentar as necessidades e dificuldades às entidades

³¹ <http://www.nomeubairro.pt/> (visitado em Agosto de 2016)

³² <http://cidadania20.com/projectos/a-minha-rua/> (visitado em Agosto de 2016)

³³ <http://cidadania20.com/projectos/mais-cidadao/> (visitado em Agosto de 2016)

responsáveis. Não é garantido que as informações reportadas cheguem efetivamente aos responsáveis e que os mesmos atuem em consonância, no entanto é uma forma de observação do espaço urbano.

Outro fator que não poderemos descurar é a constante mutação e dinâmica que existe nos espaços urbanos. As nossas cidades estão em constante transformação: novas habitações; novos centros comerciais ou zonas industriais; uma indústria ou comércio que fecha; obras ou manutenção nas infraestruturas da cidade; enfim, situações que podem originar a mudança de hábitos de mobilidade humana. O que até aí servia para colmatar as necessidades passa a ser escasso e como tal é necessário novo planeamento.

De forma a contribuir para a resposta a estas questões é proposta uma arquitetura base para o **ecossistema de mobilidade urbana** (Figura 3.5).

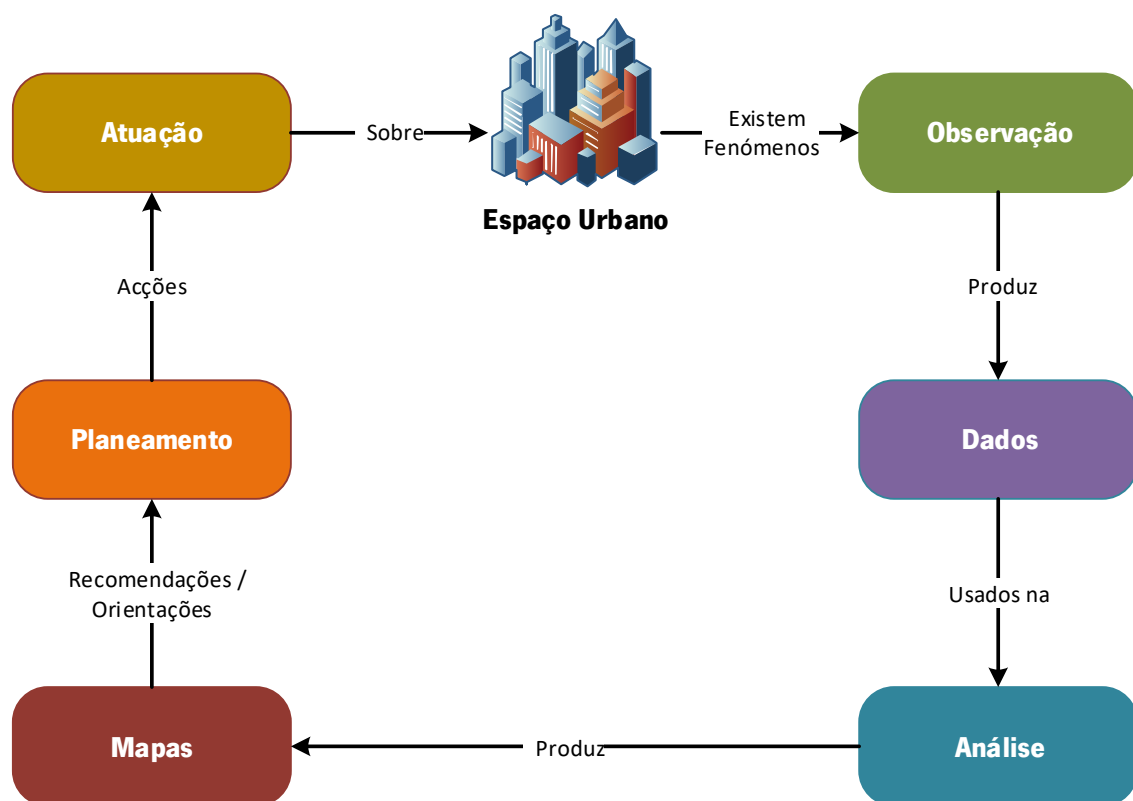


Figura 3.5 – Ecossistema de mobilidade urbana

A Figura 3.5 ilustra a proposta do que poderiam ser as entidades e interações mínimas para garantir, por um lado a contínua observação do espaço urbano, e por outro a adequada atuação de acordo com a mais recente informação disponível. Começando pelo espaço urbano, este e os indivíduos que com ele interagem estão na origem de fenómenos associados à mobilidade humana que deverão ser observados; a Observação destes fenómenos através dos mais variados sensores produz um conjunto mais ou menos alargado de Dados; por sua vez estes dados são a base da Análise

que se pretende realizar; durante a análise dos dados são produzidos Mapas de mobilidade humana, individuais e globais, dos quais surtirão recomendações ou preocupações relativamente a constrangimentos/anomalias identificados; estas recomendações permitirão um adequado Planeamento e conjunto de ações a realizar; e é com base nestes resultados que se realiza a Atuação sobre o espaço urbano.

Este processo cíclico é consecutivamente repetido uma vez que após as intervenções no espaço urbano é perfeitamente normal que sejam alterados os comportamentos e como tal, os mesmos devam agora ser identificados. Também é natural que, com o tempo e devido às dinâmicas existentes nos espaços urbanos, seja necessário com alguma periodicidade medir o estado atual da mobilidade humana.

A disponibilização e fácil utilização de dados gerados pelos dispositivos eletrónicos, que nos acompanham ou que equipam as nossas cidades, torna possível a utilização destes dados para a extração de análises sobre o estado atual da cidade, permitindo compará-los com análises anteriores, análises essas que podem quase ser feitas em tempo real, possibilitando um estudo de curto prazo sendo possível testar e avaliar o efeito de alterações no espaço urbano. Desta forma elimina-se a necessidade de realizarmos dispendiosos e morosos questionários ou censos que ao fim de algum tempo ficam desatualizados e como tal nada adequados no apoio à decisão.

A aplicação de uma arquitetura igual ou próxima à apresentada anteriormente é de enorme relevância no planeamento dos espaços urbanos pois fecha um ciclo com início na observação, passando pela decisão e culminando na atuação. O que tem sido prática comum é decidir e intervir sem o mínimo de informação necessária para a correta tomada de decisão, muitas vezes decidindo mesmo sem se conhecer as reais necessidades quer da cidade quer dos seus habitantes.

São muitos os exemplos que poderiam ser usados para justificar a razão das várias entidades incluídas nesta arquitetura, mas vejamos apenas dois exemplos para a correta contextualização:

- A mais antiga autoestrada portuguesa (A1) tinha até há poucos anos ao longo da sua extensão, em regra, duas faixas de rodagem para cada um dos sentidos. Com o aumento de tráfego o planeador decidiu pela adição de mais uma faixa em cada sentido, passando assim de um total de quatro para seis faixas.

- Mais recentemente³⁴ (Junho de 2015) foi divulgada a proposta de uma nova autoestrada que complementarà o atual Itinerário Principal 3 (IP3) que liga as cidades de Viseu e Coimbra.

Estes dois exemplos têm muito em comum, quando se identifica um aumento de tráfego e consequentemente constantes congestionamentos, a solução adotada pelos planeadores é a mais fácil e possivelmente também a mais dispendiosa para toda a população, aumentar a capacidade das infraestruturas.

Combate-se o problema não na origem mas na consequência. Como o tráfego aumentou então aumenta-se a capacidade, sem nunca se questionar o porquê deste aumento de tráfego. Porque motivo os cidadãos sentiram necessidade de usar mais estas infraestruturas? O que mudou e que originou esta necessidade? E como não existe informação que sustente os verdadeiros motivos que modificaram as necessidades dos cidadãos qualquer planeamento de extensão da capacidade será sempre mais ou menos aleatório. Por exemplo, se não sabemos qual foi o aumento de veículos que circulam no IP3 nos últimos anos, dificilmente conseguiremos planear a capacidade da nova infraestrutura a médio e longo prazo. Conclusão, esta infraestrutura poderá ficar ou subdimensionada ou sobredimensionada, com claras implicações no erário público.

Em conclusão, de forma a mudarmos o atual paradigma de solucionar algumas das questões da mobilidade humana através da construção ou reforço de infraestruturas, é fundamental decidir com base na observação e análise da mobilidade. Tal apenas será possível se tivermos um ecossistema de verdadeira cooperação entre os cidadãos, os planeadores e o espaço urbano.

³⁴ <http://www.asbeiras.pt/2015/06/governo-garante-projeto-para-construcao-de-corredor-alternativo-do-ip3-entre-coimbra-e-viseu/> (visitado em Agosto de 2016)

Capítulo 4. Estrutura de Conceitos

O ecossistema de mobilidade humana descrito no capítulo 3 necessita para a sua adequada implementação e correto funcionamento, de uma participação efetiva tanto dos decisores como dos cidadãos em dois momentos: observação e atuação. Estes dois atores são fundamentais para a contínua avaliação do espaço urbano e respetiva atuação. De nada serve observar se não se intervém, assim como não se deve intervir se não conhecermos a verdadeira realidade dos problemas ou necessidades.

Dada a importância da observação da mobilidade humana e tendo em consideração que o leque de sensores que permitem essa observação é cada vez mais alargado, é relevante que se faça uma representação uniforme dos dados recolhidos através da observação e que os dados de diferentes fontes e formatos possam ser representados segundo a mesma estrutura facilitando posteriormente a análise e a visualização da informação disponibilizada.

4.1. Estrutura de Conceitos

Com o objetivo de contribuir para a uniformização da estrutura de informação associada à mobilidade humana, ao longo desta secção serão apresentados os conceitos e processos de transformação dos dados que caracterizam a estrutura de conceitos desenvolvida. Estes conceitos foram pensados e desenhados para se adequarem aos dados desde a fase de aquisição até à representação homogénea de informação relativa à mobilidade humana de um único indivíduo ou de um grupo de indivíduos. Os nomes dos conceitos e atributos que os caracterizam foram mantidos em inglês de forma a ser mais coerente com outros trabalhos da área existentes na literatura.

4.1.1. Dados em bruto

A base da observação da mobilidade humana está nos dados recolhidos a partir de um conjunto distinto de sensores, dados estes obtidos em bruto diretamente dos sensores e sem qualquer processamento relevante, conhecidos como **raw data**, que refletem a movimentação de um artefacto ao longo do tempo. Até há algum tempo atrás os recetores GPS eram a principal fonte de dados associados ao movimento. No entanto, com os recentes avanços tecnológicos, em particular nos dispositivos móveis, surgiu uma nova oportunidade de recolhermos quantidades avultadas de dados acerca da mobilidade dos utilizadores. Estes dispositivos, para além de recolherem dados georreferenciados através dos recetores GPS (a Tabela 4.1 ilustra um exemplo de dados disponíveis) que se encontram atualmente integrados em toda a gama de smartphones, permitem também a recolha de dados relativos ao uso das redes Wi-Fi (Tabela 4.2) e GSM (Tabela 4.3), deteção de dispositivos próximos (indivíduos), registo de dados gerados pelos acelerómetros e outros dados oriundos dos mais variados sensores que equipam atualmente estes dispositivos.

Timestamp	Latitude	Longitude	Altitude	Velocidade
2011/06/29 15:25:07	1,297077	103,7808	93,5	0,75
2011/06/29 15:25:18	1,297077	103,7808	108,2	0,75
2011/06/29 15:25:31	1,297213	103,7806	134,4	1

Tabela 4.1 – Dados em bruto do sensor GPS

Timestamp	BSSID	RSSI	SSID
2011/06/29 15:25:08	00:27:0d:07:d6:c0	-90	NUS
2011/06/29 15:25:11	00:27:0d:07:d6:c0	-88	NUS
2011/06/29 15:25:12	00:27:0d:07:d6:c0	-88	NUS

Tabela 4.2 – Dados em bruto do sensor Wi-Fi

Timestamp	CID	LAC	MNC	SIGNAL_STRENGTH
2011/06/29 15:25:08	962335	441	3	9
2011/06/29 15:25:10	962335	441	3	8
2011/06/29 15:25:11	962335	441	3	8

Tabela 4.3 – Dados em bruto do sensor GSM

Mas os dados de mobilidade não se restringem apenas aos que podemos e conseguimos recolher dos smartphones. As infraestruturas urbanas também têm contribuído para o aumento das

capacidades sensoriais. Por exemplo, os operadores de transporte público podem tirar partido dos dados recolhidos dos sistemas de bilhética (registos de entradas e saídas dos passageiros) quer sejam autocarros ou metro, de forma a adaptar a oferta às necessidades dos cidadãos. As autoridades públicas podem também, por exemplo, recolher dados de intensidade de tráfego em determinadas artérias de uma cidade.

Quer seja então através dos dispositivos móveis ou de infraestruturas urbanas, entre outros exemplos, os ambientes urbanos estão atualmente tecnologicamente equipados para, do ponto de vista da análise de mobilidade humana, serem uma rede de sensores gigante produtora de elevada quantidade de dados.

O desafio particular dos conjuntos de dados em bruto recolhidos no ambiente urbano prende-se com o facto de os mesmos serem heterogéneos: quer no formato usado para a representação dos dados de mobilidade; quer na resolução e precisão da deteção da posição do artefacto; quer na precisão do registo temporal; quer na frequência com que os dados são recolhidos; e por último, no referencial espacial que é usado para a representação dos pontos ou locais no espaço.

A heterogeneidade do referencial espacial é de grande relevância quando pretendemos relacionar ou associar dados oriundos de diferentes tipos de sensores (fusão de dados). Vejamos o seguinte exemplo, meramente ilustrativo para a descrição da localização da Universidade do Minho em Guimarães, usando dois dos maiores referenciais espaciais:

1. Universidade do Minho, Campus de Azurém, 4800-058 Guimarães
2. Latitude: 41.4521216; Longitude: -8.2913494

No primeiro temos uma descrição da localização baseada no nome e endereço do local, no segundo uma descrição de acordo com coordenadas GPS no referencial WGS-84 (National Imagery and Mapping Agency (NIMA), 2000). Além da distinção relativamente ao referencial utilizado para descrever o local, o mais significativo ainda é o modelo de representação usado. Enquanto no primeiro caso temos um modelo de representação simbólica, no segundo é um modelo de representação geométrica. Relativamente a este modelo geométrico, é o mais vulgarmente usado, pelo que não levanta em si grandes dificuldades; contrariamente ao modelo simbólico onde a diversidade é grande: endereço postal, código postal, identificadores de redes celulares, identificadores de paragens de autocarros e metro, entre outros.

De forma a ser viável agregar ao máximo esta heterogeneidade dos dados num modelo o mais genérico e homogéneo possível foi elaborado um conjunto de conceitos para a representação da

informação de mobilidade. As subsecções seguintes descrevem os vários conceitos que foram identificados e respetiva fundamentação.

4.1.2. Observação

Um artefacto móvel pode ser observado de diferentes perspetivas através de diferentes sensores. Cada um desses sensores pode recolher dados sobre os artefactos com diferentes atributos e funcionalidades. Como os dados em bruto incluem diferentes atributos é necessária uma forma de os normalizar.

Com base nesta necessidade um dos primeiros conceitos associados à mobilidade é o conceito de ***Observation***. Este conceito tem assim como objetivo a realização de um primeiro passo na normalização, definindo um conjunto básico de informação necessária para caracterizar onde e quando foi observado o artefacto a partir do ponto de vista da mobilidade. Independentemente do sensor utilizado para a aquisição dos dados de mobilidade em bruto, estes dados refletem a observação de um fenómeno. A observação de um único artefacto pode ser descrita de forma abstrata da seguinte forma:

A presença de um artefacto num ponto específico do referencial espaçotemporal

e representado pelos seguintes atributos:

(Id_ Observation, Artefact, Location, Timestamp)

O primeiro atributo permite a distinção entre as observações do mesmo artefacto no mesmo local e tempo quando realizada por múltiplos sensores. O atributo *Artefact* identifica o artefacto que foi observado, facilitando a criação de vistas individuais. O *Timestamp* regista o momento no tempo em que a observação se realizou. O atributo *Location* descreve, com diferentes níveis de detalhe, a localização do artefacto no momento em que foi observado. Como os dados em bruto obtidos de diferentes sensores podem definir a localização de um artefacto de acordo com referenciais simbólicos ou geométricos, o atributo *Location* combina assim ambas as representações: geométrica, quando a representação da localização é baseada em coordenadas cartográficas, e simbólica, quando a localização é caracterizada pela sua própria descrição, por exemplo, nome do edifício, uma morada, um identificador de uma célula da rede GSM, entre outros.

Relativamente à dimensão temporal de uma observação existe um referencial e uma definição quase universal da mesma, pelo que a definição do atributo *Timestamp* numa observação não causa

grandes dificuldades. Deveremos ter alguns cuidados relativamente aos fusos horários, por exemplo, e aspetos relacionados com sincronização deste referencial ao longo dos diferentes sensores usados. Lidando de forma consciente com estas questões, a dimensão temporal não levanta assim grandes preocupações.

No entanto, o mesmo não se verifica com a dimensão espacial, onde esta uniformidade não existe, uma vez que o espaço pode ser descrito de forma geométrica ou simbólica e consequentemente utilizar diferentes e independentes referenciais (Figura 4.1b).

4.1.2.1. Localização

Independentemente do tipo de sensor e consequentes dados recolhidos, usados para a obtenção de uma observação, cada registo de observação descreve um artefacto como um ponto num referencial espaçotemporal (como apresentado na Figura 4.1a). Um artefacto é assim descrito segundo duas dimensões: tempo e espaço.

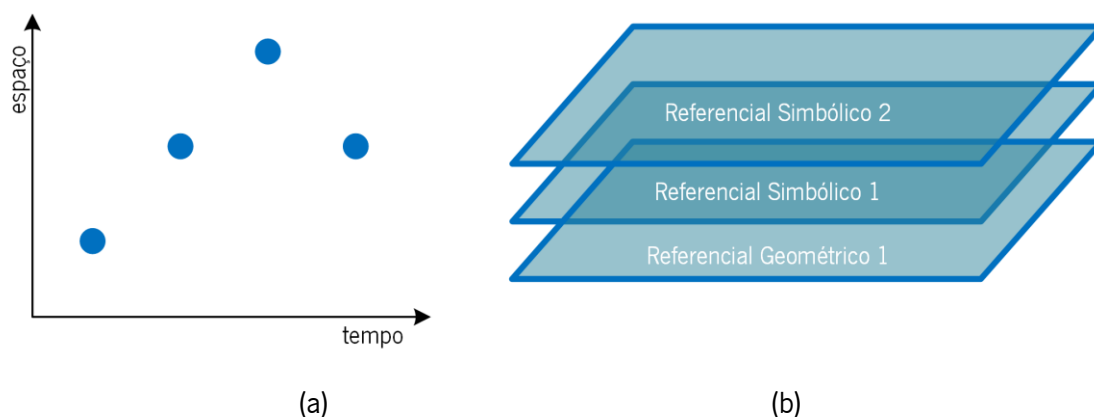


Figura 4.1 – Observações num referencial espaçotemporal (a) e a multiplicidade das dimensões espaciais (b)

Com o objetivo de normalizar diferentes referenciais numa observação, foi necessário definir o conceito **location** como um ponto num referencial espacial arbitrário, seja ele geométrico e/ou simbólico, e representado da seguinte forma:

$$\text{location} = \{\text{symbolic name}, \text{position}\}$$

onde o atributo *position* é representado por um par de coordenadas na forma:

$$\text{position} = (\text{latitude}, \text{longitude}) \text{ ou } \text{position} = (X, Y)$$

Além de serem representados de forma diferente, os referenciais simbólicos e geométricos têm ainda diferentes características no que se refere à análise de movimento. Enquanto os espaços

geométricos são regra geral contínuos, os espaços simbólicos são usualmente discretos. Inclusivamente, a noção de distância num referencial geométrico é de simples cálculo recorrendo a uma de várias funções de distância: Euclidiana, Manhattan, entre outras; enquanto o cálculo de distâncias em domínios simbólicos não é direto.

4.1.3. Lugar

Na literatura são usados vários termos para descrever a dimensão espacial no contexto de análise de mobilidade: posição, localização, local, lugar, entre outros. Uma *location* como definida anteriormente, não é em si muito relevante para a caracterização da mobilidade de um indivíduo específico, uma vez que apenas representa um ponto num referencial espacial, geométrico ou simbólico. O que é relevante para a caracterização de um indivíduo, do ponto de vista da mobilidade, é o **Place**. O conceito de *Place* está muito mais próximo do conceito mental de uma região ou área num espaço real com efetivo significado para um indivíduo: a casa, o local de trabalho, uma praça ou parque, um centro comercial, ou até mesmo uma cidade. Um *Place* é assim uma região no espaço real e nos modelos mentais humanos de espaço, região esta com uma fronteira por vezes indefinida e caracterizada por ser relevante ou a um indivíduo ou a um grupo de indivíduos.

No contexto da estrutura de conceitos que foi desenvolvida, um *Place* é definido através da agregação de *Observations* próximas, e pode este conceito ser descrito de forma abstrata da seguinte forma:

Conjunto de Observações que caracterizam uma região relevante

e representado pelos seguintes atributos:

(Id_Place, List of Observations)

Um *Place* deixa de ser apenas um ponto específico num referencial espacial, mas antes uma região no espaço definida pela agregação de um conjunto de *Locations* que fazem parte das *Observations*, independentemente de as localizações serem simbólicas ou geométricas. Adicionalmente a dimensão temporal das *Observations* agregadas definem o perfil temporal de um *Place*, contrariamente a uma única *Observation* que é apenas uma representação instantânea da presença de um artefacto algures no tempo e espaço.

A necessidade de agregarmos as *Locations* pode ser demonstrada pela seguinte realidade. Considerando que a localização de um artefacto numa sequência de observações é descrita de forma geométrica (por exemplo, através de pares de coordenadas GPS), um artefacto imóvel pode reportar

diferentes pontos no domínio espacial, ao longo de consecutivas observações, devido ao erro/ruído no sensor. Cada um destes pontos (localizações) individualmente não têm qualquer significado para a caracterização da mobilidade, pois mesmo diferentes, nós sabemos que o artefacto não se moveu. Um fenómeno semelhante pode ser observado num domínio de representação simbólica da localização, por exemplo, o artefacto poderá estar igualmente imóvel mas a informação extraída por exemplo de um sensor Wi-Fi pode estar a reportar observações consecutivas com diferentes identificadores de pontos de acesso.

A agregação de localizações não é tão significativa nos referenciais simbólicos, uma vez que estes são normalmente discretos, contrariamente aos referenciais geométricos que, sendo regra geral contínuos, são mais sensíveis ao ruído nos sensores e como tal onde a agregação se torna mais relevante. Por exemplo, se tivermos, mesmo que pequenas, alterações na posição do artefacto, ao realizarmos o processo de agregação e consequente deteção e extração do *Place*, obtemos uma melhor caracterização geográfica do artefacto e consequentemente um número menor de lugares distintos visitados pelos artefactos.

A Figura 4.2 ilustra o conceito de *Place* como uma agregação de várias *Observations*. A área identificada na figura composta por várias observações circunscritas pelo contorno, define um *Place* extraído e os círculos são a representação da *Location* das *Observations*. A seleção de quais as localizações que pertencem ou não a um determinado lugar baseia-se em critérios de proximidade.

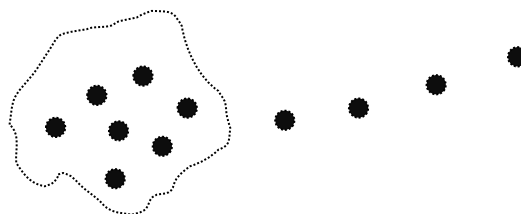


Figura 4.2 – Extração de um Lugar

Extrair *Places* a partir de um conjunto de *Observations*, uma atividade conhecida como *place learning*, requiere diferentes processos para lidar com dados geométricos e com dados simbólicos. Além disso, os processos de agregação estão também dependentes da aplicação específica da informação, uma vez que cada área de aplicação requiere diferentes níveis de detalhe na informação. Por exemplo, considerando a análise da mobilidade humana do ponto de vista dos transportes, de nada serve tentarmos distinguir diferentes lugares dentro de um edifício de pequenas dimensões. As movimentações dentro de um edifício podem ser relevantes para caracterizar o perfil de mobilidade individual, mas para a área aplicacional dos transportes é irrelevante uma vez que são trajetos

realizados sem a necessidade de meios de transporte. Por outro lado, para sistemas de navegação dentro de edifícios, a distinção entre diferentes divisões ou pisos no mesmo edifício são de enorme importância. Por estes motivos, pode ser necessário definir e aplicar diferentes critérios de agregação no momento de decisão pela adição ou não de uma observação a um determinado lugar.

A maioria dos processos de *place learning* descritos na literatura são baseados em variantes dos algoritmos de *clustering*. Um desses exemplos é o trabalho de Andrienko (Andrienko *et al.*, 2011). Quando aplicados a um conjunto de *Locations*, estes processos classificam cada uma das localizações como parte de um *cluster* ou como ruído (*outliers*). Idealmente cada um dos *clusters* obtidos deverá corresponder a um dos locais visitados pelo artefacto que está a ser observado, e os elementos identificados como ruído corresponderem a observações obtidas durante a movimentação do artefacto. Na prática, mesmo os melhores algoritmos de *place learning* tendem a detetar mais locais do que aqueles reconhecidos pelo artefacto observado. Um dos motivos deve-se ao facto destes algoritmos terem que estabelecer um compromisso entre as taxas de falsos positivos e falsos negativos. Outro motivo é o facto da maioria dos indivíduos não reconhecerem como relevante alguns dos locais que foram detetados. Exemplos desses casos são as paragens em semáforos, ou locais que foram visitados apenas uma vez. Uma possível solução para esta questão poderá ser assumir como “candidatos” ou “potenciais” todos os lugares identificados pelos *clusters* e destes extrair um subconjunto de lugares que são efetivamente “relevantes” ou “significativos” para o indivíduo ou para um grupo de indivíduos. A classificação e distinção entre locais relevantes e locais candidatos pode ser feita com base num conjunto de métricas, como por exemplo a quantidade de tempo passado em cada um dos locais ou então a frequência com que os locais são visitados. Mais uma vez, as métricas e critérios a serem adotados para classificar os locais como relevante ou não está dependente do domínio aplicacional. Como exemplo, consideremos o caso de um indivíduo que viaja de carro para um local distante todos os meses e a sua estadia nesse local é de apenas uns minutos. Este local não é relevante para o indivíduo nem devido ao tempo total de presença no local, nem pela frequência das visitas. No entanto, se considerarmos que o local fica longe de sua casa, já poderá ser relevante uma vez que os custos de viagem são elevados. Este exemplo demonstra que, em alguns casos, a relevância de um local não pode ser definida apenas a partir das características intrínsecas do local, mas devem também incluir o relacionamento com outros lugares e viagens realizadas entre eles.

A relevância dos locais pode ainda ser diferente se considerarmos a perspetiva pessoal, coletiva ou espacial. A nossa casa é certamente um dos locais mais relevantes na perspetiva pessoal, mas

cada um desses locais, individualmente, não é relevante para a população de toda uma cidade. Por outro lado, um bar visitado ocasionalmente por um indivíduo pode não ser nada relevante para este na perspectiva pessoal, mas poderá ser relevante, por exemplo, para a análise de propagação de doenças, uma vez que representa um local onde um conjunto alargado de indivíduos se concentra. Existem outros locais que são apenas relevantes na perspectiva da organização e planeamento do espaço. Um desses exemplos pode ser uma rotunda onde um conjunto de veículos converge em determinado período temporal do dia. O facto de um local não ser relevante para o indivíduo ou grupo de indivíduos, pode ser relevante para o planeamento de tráfego de uma cidade. Neste contexto, os locais candidatos de um indivíduo podem, após agregação, passar a ser locais relevantes numa perspectiva coletiva ou espacial.

Até ao momento vimos o *Place* como uma agregação de um conjunto de *Locations* extraídos a partir das *Observations*. No entanto vimos anteriormente que podemos ter múltiplos sensores e como tal múltiplas observações de um indivíduo em simultâneo, e muito provavelmente com referenciais espaciais diferentes. Neste caso a agregação não poderá ser feita com base na proximidade espacial, uma vez que não é possível associar diretamente localizações com características diferentes. Por exemplo, se tivermos uma observação através de um sensor GSM e uma observação de um sensor GPS, é totalmente impossível medir a distância espacial entre estas duas observações uma vez que temos uma localização simbólica e outra geométrica. Mesmo que as localizações sejam simbólicas, por exemplo de um sensor GSM e de um sensor Wi-Fi, como estas usam referenciais diferentes não é possível estabelecer uma proximidade entre as observações. Por este motivo, a deteção de *Places* poderá ser, neste caso, feita não com base na proximidade espacial, mas sim na proximidade temporal (Hightower, Consolvo e LaMarca, 2005).

Da mesma forma que várias *Locations* extraídas de *Observations* consecutivas, quando agregadas, dão origem a um *Place* (com base na distância ou critérios de proximidade), o mesmo se aplica à agregação de *Observations* oriundas do mesmo sensor ou de sensores diferentes com base no seu atributo *Timestamp*. Assim a agregação é feita com base em critérios de proximidade temporal e um *Place* passa a ser descrito por múltiplas *Locations* mesmo que estas tenham representações simbólicas e geométricas.

4.1.4. Suspensão do movimento

Analisando um conjunto de observações consecutivas $\{..., O_{i-1}, O_i, ...\}$ de um artefacto ao longo do tempo, este artefacto pode ser classificado como estando em um de dois estados possíveis: **em movimento** ou **imóvel**. Um artefacto está no estado imóvel quando, nas observações mais recentes e consecutivas, o *Place* (P) é o mesmo, independentemente do valor presente no atributo *Location* (l):

$$(l_i \in P_j) \wedge (l_{i-1} \in P_j) \quad (1)$$

Se, pelo contrário, tivermos observações consecutivas a que correspondem *Places* diferentes ou inexistentes, então o nosso artefacto está em movimento.

A presença de um artefacto durante um período de tempo no mesmo *Place* permite definir um novo conceito de **Stay**, ou estadia, que de forma abstrata poderá ser descrito como:

Intervalo de tempo entre a primeira e última Observação de um artefacto no mesmo Lugar

e pode ser representada pelos seguintes atributos:

(Id_Stay, Artefact, Place, Timestamp_Initial, Timestamp_Final)

Uma *Stay* ocorre assim quando para um conjunto de observações consecutivas do mesmo artefacto o *Place* mantém-se o mesmo. O número e duração das *Stays* depende quer dos *Places* que foram identificados, e consequentemente, depende também da forma como as *Observations* são agregadas. No caso de representações simbólicas dos *Places*, o processo de transformação que permitem a extração das *Stays* a partir das *Observations* é simples, contrariamente às representações geométricas, devido à flutuação da informação relativo ao posicionamento do artefacto (ruído). Vários autores deram particular atenção ao problema da detecção de *Stays* nos *Places* (por vezes apelidados de *stops* ou regiões de suspensão de movimento), em particular para os casos em que temos informação GPS. O conceito aqui apresentado e a problemática associada à extração de *Stays* está orientada não apenas para informação de GPS, mas para múltiplos tipos de dados oriundos de múltiplos tipos de sensores.

A definição de uma *Stay* tem, no entanto, duas restrições formais. Primeiro, o intervalo de tempo deverá ser sempre maior que zero, de outra forma não seria uma estadia, mas uma observação instantânea possivelmente originada pela movimentação do artefacto. Segundo, o maior intervalo de tempo entre duas *Observations* consecutivas (a partir das quais a *Stay* emerge) deve ser curto o

suficiente de forma que se possa assumir que o artefacto não abandonou, entretanto, o correspondente *Place* entre *Observations*. De outra forma, se o intervalo entre as *Observations* for longo, não poderemos extrair uma *Stay* uma vez que não é razoável inferir observações intermédias (ao longo de uma linha reta que ligue duas observações no referencial espaçotemporal). Este tipo de situações ocorre, por exemplo, quando um artefacto entra dentro de um edifício e a partir desse instante perdemos o seu posicionamento através do GPS. A ausência de observações não significa que o artefacto tenha abandonado o edifício. De forma a darmos resposta a este tipo de situações foi definido um novo conceito, ***Time Leap***, que pode ser descrito da seguinte forma:

Longo período de tempo entre duas Observações consecutivas do mesmo artefacto no mesmo Lugar

e representado pelos seguintes atributos:

(Id_TimeLeap, Artefact, Place, Timestamp_Initial, Timestamp_Final)

Quando comparada a representação do conceito *Time Leap* com a representação do conceito *Stay* verificamos que são formalmente idênticas, no entanto com semânticas distintas. No caso de uma *Stay* temos observações que permitem sustentar que o artefacto se manteve efetivamente no mesmo *Place*. No caso de um *Time Leap* pela ausência de observações apenas poderemos afirmar que o artefacto foi identificado num *Place* e mais tarde voltou a ser identificado no mesmo *Place*, mas por falta de observações durante esse período é impossível saber se o artefacto se movimentou ou não. De forma abstrata, poderemos afirmar que o conceito *Time Leap* corresponde a um “salto temporal” dentro do mesmo *Place* (Figura 4.3).

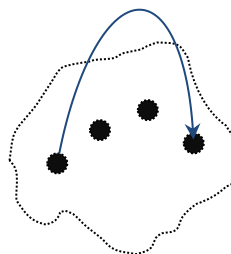


Figura 4.3 – Um *Place* com quatro *Observations* diferentes e um *TimeLeap* entre duas delas

4.1.5. Movimento

Os conceitos *Stay* e *Time Leap* focam situações em que os artefactos se encontram no estado imóvel, mesmo que existam ligeiros movimentos, mas irrelevantes na perspetiva da mobilidade. No entanto, a origem de alguns dos problemas ou dificuldades nas cidades está precisamente na

mobilidade dos indivíduos. Assim um artefacto diz-se que se encontra em movimento quando existe uma mudança de *Location* (l) entre duas *Observations* consecutivas, e estas não estão associadas ao mesmo *Place* (P) (uma ou ambas as observações podem inclusivamente não estar associadas a nenhum *Place* em particular):

$$(l_i \neq l_{i-1}) \wedge [\forall_j (l_i, l_{i-1}) \notin P_j] \quad (2)$$

Assim quando um artefacto se encontra em movimento, esperamos observar variações no atributo *Location* ao longo do tempo. Tendo em conta esta constatação é necessário termos um conceito que represente esta variação, o ***Elementary Movement***, descrito da seguinte forma:

A mudança da Localização de um artefacto observada entre Observações consecutivas

e representada pelos seguintes atributos:

(Id_Movement, Artefact, Location_Start, Location_End, Timestap_Initial, Timestap_Final)

Este conceito tem uma amplitude muito restrita, pelo que não se considera qualquer mudança de *Location* como um *Elementary Movement*. Apenas se pode considerar que um *Elementary Movement* existe quando para um par de *Observations* consecutivas com diferentes valores para o atributo *Location* é razoável inferir (por interpolação) uma posição intermédia no intervalo de tempo entre as duas observações. Apenas em situações onde podemos assumir que o artefacto se moveu de uma *Location* inicial para uma *Location* final ao longo de uma linha reta a uma velocidade constante, ou que o movimento real é razoavelmente assim representável. O melhor exemplo desta situação é quando transformamos um conjunto de observações consecutivas obtidas de um recetor de GPS numa linha representativa da trajetória realizada pelo artefacto. Se o artefacto que estamos a observar realiza um percurso pedestre e adquirimos dados de GPS com uma frequência de um segundo é bastante razoável inferir que o artefacto realizou uma sequência de *Elementary Movements*, por ser razoável inferir posições intermédias entre duas observações consecutivas no tempo. Agora, se tivermos uma outra situação onde o intervalo de tempo entre observações consecutivas for na ordem de um minuto, não é adequado afirmar que estamos perante *Elementary Movements*, uma vez que o espaço percorrido neste intervalo de tempo poderá ser bastante significativo, não permitindo a inferência de posições intermédias entre as duas localizações (inicial e final).

O conceito *Elementary Movement* deve ser usado com algum cuidado e de acordo com o cenário aplicacional. Por exemplo, se para um percurso pedestre ter observações com intervalos de tempo de um minuto não permite de forma razoável inferir este tipo de movimentos, por outro lado, se as observações se referirem a um voo de um avião, o intervalo de tempo de um minuto poderá ser um valor adequado para extrairmos *Elementary Movements*. Assim os critérios associados à inferência deste conceito deverão ter em conta as tipologias dos dados e áreas aplicacionais.

A Figura 4.4 ilustra o conceito de *Elementary Movement*, neste caso como uma sequência de setas que interligam entre si *Observations* consecutivas.

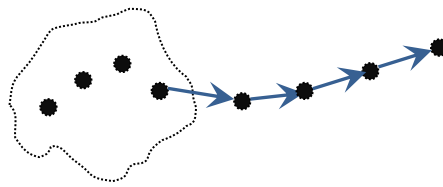


Figura 4.4 – Sequência de *Elementary Movements* entre *Locations*

De salientar, pela definição, que os *Elementary Movements* não podem ser extraídos a partir de um par de *Observations* em que ambas são membros do mesmo *Place*. No entanto, se for apenas uma *Observation* a fazer parte do *Place*, nesse caso pode ser usada para definir um *Elementary Movement*.

Como vimos na definição do conceito, em momentos onde o intervalo de tempo entre observações consecutivas de um artefacto não permite inferir as posições intermédias devido ao grande espaço de possibilidades, então não poderemos afirmar que um *Elementary Movement* ocorreu. No entanto poderemos definir um novo conceito, o ***Space Leap***. O conceito de *Space Leap* pode ser descrito como:

Uma mudança da Localização de um artefacto observada entre Observações consecutivas

e poderemos representá-lo pelos seguintes atributos:

(Id_SpaceLeap, Artefact, Location_Start, Location_End, Timestamp_Initial, Timestamp_Final)

Como acontecia com as representações de *Stay* e *Time Leap* também as representações dos conceitos *Space Leap* e *Elementary Movement* são similares na forma, embora, mais uma vez, com semântica diferente. A Figura 4.5 ilustra de forma abstrata o conceito de *Space Leap*, neste caso como

um “salto” entre dois locais diferentes. De modo geral, um *Space Leap* ocorre entre duas *Observations* consecutivas desde que as mesmas não estejam associadas ao mesmo *Place*.

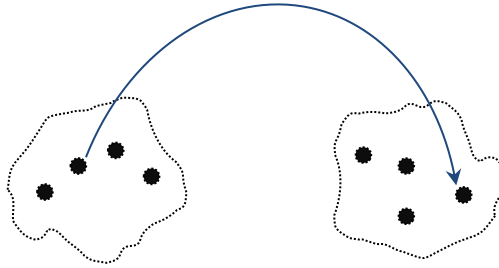


Figura 4.5 – Um exemplo de um *Space Leap*

4.1.6. Trajetória

O conceito ***Trajectory*** é, provavelmente, o conceito mais complexo usado na análise de mobilidade. Muitas vezes, a trajetória é vista como o caminho percorrido ao longo do tempo. Esta definição encaixa bem com a natureza contínua, tanto nos domínios do espaço e do tempo no mundo real, mas não tão bem com a natureza da observação tecnológica, discreta, do artefacto observado. Na grande maioria das situações, as tarefas de análise de movimento são feitas por manipulação/tratamento de dados gravados digitalmente. Os registos de dados correspondentes, que representam pontos do referencial espaçotemporal (que foram apelidadas de *Observations*) são, frequentemente, obtidos em intervalos irregulares no tempo ou no espaço, ou em ambos.

Além disso, enquanto se observa o movimento de indivíduos durante longos períodos de tempo, podem ocorrer falhas temporais no processo de observação. Como consequência, o conceito de trajetória no contexto de análise de movimento deve acomodar essas imperfeições do processo de observação, de modo a ser possível extrair conclusões significativas a partir de dados reais.

A definição do conceito de *Trajectory* foi baseada na ideia de alternância entre os estados de movimento e suspensão de movimento. Uma vez que a suspensão de movimento está associada a *Stays* e *Time Leaps* em *Places*, qualquer outro evento entre o abandonar de um *Place* e o entrar no mesmo ou em outro *Place* é considerado então como uma *Trajectory*. O conceito é então descrito da seguinte forma:

Lista ordenada no tempo de Movimentos Elementares e Saltos Espaciais que ocorrem entre a saída de um Lugar e a entrada num Lugar

e representado pelos seguintes atributos:

(Id_Trajectory, Artefact, Place_Start, Place_End, List of Elementary Movements and Space Leaps)

A Figura 4.6 ilustra o conceito de *Trajectory*, descrevendo os três diferentes tipos de trajetórias: (i) descrito por uma lista de *Elementary Movements*; (ii) descrito por uma lista de *Space Leaps*; (iii) e por último, descrito por uma mistura de *Elementary Movements* e *Space Leaps*. De salientar que o exemplo ilustrado para o segundo caso, confunde-se com o conceito de par origem-destino (onde a origem define o ponto de partida de uma deslocação e o destino o ponto de chegada).

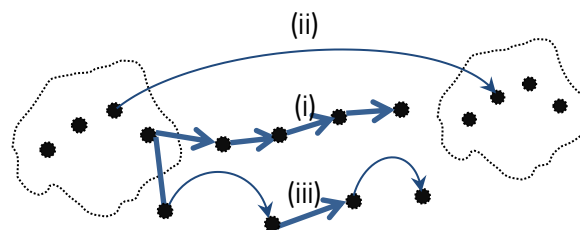


Figura 4.6 – O conceito de *Trajectory*

No trabalho apresentado por Brockmann (Brockmann e Theis, 2008), a representação de trajetórias recorrendo a pares origem-destino é provavelmente a abordagem mais apropriada dado o problema que estava a ser abordado. Uma vez que, neste tipo de cenário de mobilidade urbana, o objetivo é descobrir de onde e para onde as pessoas se movem, os pares origem-destino fornecem uma representação conveniente do movimento. Uma abordagem semelhante pode ser usada em outros cenários aplicacionais, tais como o suporte para *carpooling*. Tendo em consideração que, neste cenário, o objetivo é identificar a origem e destino das viagens dos indivíduos, para avaliar em que medida eles podem partilhar o mesmo carro, a descrição das trajetórias através desta abordagem é perfeitamente adequada. No entanto, existem outros cenários de mobilidade onde representar o movimento através da utilização de pares origem-destino não é suficiente, e em que o trajeto percorrido também é importante. As áreas de seguros, ecoeficiência e planeamento de tráfego são apenas alguns desses exemplos. Para este tipo de cenários, descrever as trajetórias apenas através de pares origem-destino não ilustram os comportamentos de mobilidade dos cidadãos ou dos verdadeiros e reais problemas que existem no espaço urbano e que afetam significativamente o espaço e as pessoas que interagem com ele.

Por existirem diferentes perspectivas relativamente à noção de trajetória e o que a caracteriza, houve a preocupação de que a definição do conceito *Trajectory* abrangesse ao máximo os diferentes pontos de vista. Quando disponível, as observações obtidas através de métodos de amostragem de alta frequência podem ser usadas para representar as trajetórias utilizando sequências de *Elementary Movements*, obtendo assim resultados com abundância de detalhe espacial e temporal. Além disso, estas trajetórias detalhadas podem ser simplificadas para um conjunto mais pequeno de *Space Leaps*, ou um único par origem-destino, se a representação for suficiente para a aplicação prevista.

4.1.7. Processos de transformação

Como resultado final da definição dos vários conceitos que compõem a estrutura de conceitos desenvolvida, foi desenhada uma representação em camadas com os diferentes conceitos complementada com os necessários processos de transformação dos dados entre os diferentes conceitos.

A Figura 4.7 ilustra assim em camadas a estrutura de conceitos desenvolvida:

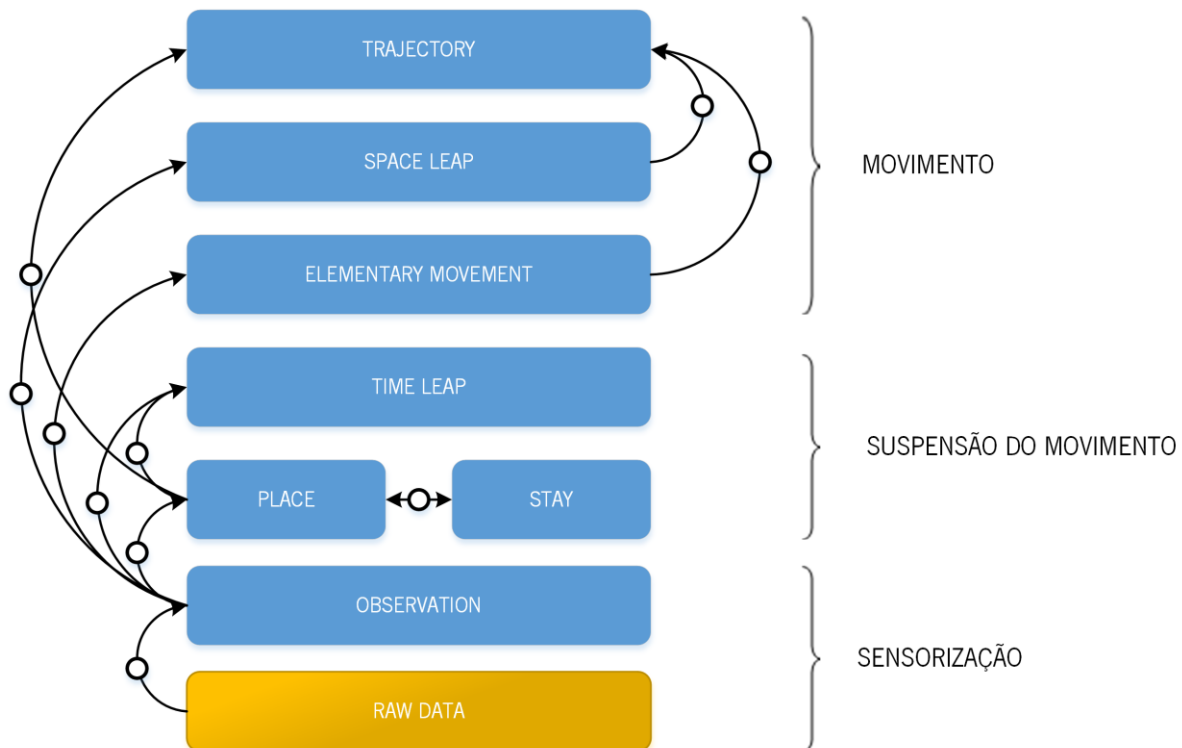


Figura 4.7 – Estrutura em camadas dos conceitos de informação

A camada associada aos *raw data* não encerra em si um conceito e por esse motivo é apresentada de cor diferente. No entanto como é a base de todo o processo evolutivo da informação de mobilidade

seria sempre fundamental a sua inclusão. Lateralmente aos conceitos, aparecem ilustrados, através de setas direcionadas, os vários processos de transformação de dados necessários para adequar os dados disponíveis nos conceitos que pretendemos atingir. Do lado direito da figura foi adicionada uma classificação dos vários conceitos de acordo com o seu enquadramento, como exemplo, o conceito *Observation* está intimamente associado aos dados obtidos dos sensores. Os restantes conceitos foram subdivididos de acordo com o estado do artefacto, se são conceitos relativos ao movimento ou à suspensão de movimento.

Como referido anteriormente na definição do conceito de *Place* existe muitas vezes necessidade de classificar os *Places* como relevantes ou não para os indivíduos ou grupos de indivíduos. Como tal foi necessário aprofundar a representação gráfica da estrutura de conceitos de forma a contemplar por um lado os locais que são identificados numa primeira fase e que são candidatos a *Place* e por outro, os locais que são efetivamente relevantes tendo em conta o cenário aplicacional em causa. Assim a Figura 4.8 ilustra com maior detalhe os dois conceitos associados ao *Place* e processos de transformação.

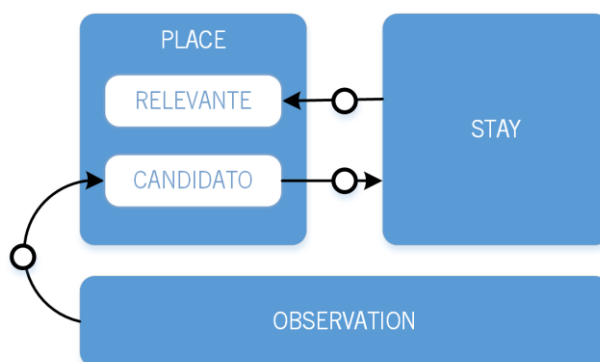


Figura 4.8 – Conceito *Place* detalhado

Todos os *Places* extraídos inicialmente a partir das *Observations* são considerados como candidatos. Posteriormente e com base nos locais que foram extraídos são geradas as *Stays* para cada um desses locais, onde consta o tempo de estadia em cada um dos locais candidatos. De seguida e com base nos critérios de elevação de um *Place* de candidato a relevante, que poderá ser por exemplo, o tempo total da estadia ou o número de ocorrências desse local (equivalente ao número de visitas), os *Places* ficam assim classificados como relevantes ou candidatos.

Capítulo 5. Experimentação

5.1. Introdução

A estrutura de conceitos desenvolvida e apresentada no capítulo anterior visa primordialmente a integração de diferentes tipologias de dados de mobilidade e adequar os mesmos de acordo com os conceitos definidos. Homogeneizar a informação de mobilidade de acordo com uma estrutura que possibilite múltiplos sensores, necessita além dos conceitos, de processos de transformação dos dados, dados estes que fluem ao longo desses mesmos conceitos.

Ao longo deste capítulo pretende-se apresentar dois exemplos da implementação dos processos de transformação e algoritmos associados a estes, e respetivos *outputs*, de acordo com dados de diferentes sensores e diferentes utilizadores, de forma a demonstrar a aplicabilidade dos conceitos de forma individual e da estrutura de conceitos como um todo para a representação de dados espacotemporais.

As duas seções seguintes permitem ainda ilustrar as vantagens e desvantagens, bem como as potencialidades, da análise da mobilidade humana quer através de dados recolhidos de forma individual, quer com base em dados extraídos de redes sociais.

5.2. Dados individuais – múltiplos sensores

A primeira experimentação realizada para validação e aperfeiçoamento da estrutura de conceitos proposta teve por base dados recolhidos de forma individual para um conjunto de diferentes utilizadores (foram recolhidos dados de cinco utilizadores diferentes com duração que variou entre um mês e três meses consecutivos). Nas próximas subsecções estão descritas as várias fases de

implementação e respetivos processos de transformação dos dados, assim como potenciais resultados que se conseguem atingir com recurso à análise de dados individuais de mobilidade. A realização de análises individuais e consequente geração de mapas individuais de mobilidade permitem demonstrar na perfeição a verdadeira noção de mobilidade, um misto de movimento e paragens. O recurso a dados agregados ao invés de dados individuais apenas permite analisar parcialmente a mobilidade, uma vez que permitem no máximo a descrição espaçotemporal da ocupação do espaço associado às paragens.

5.2.1. Aquisição dos Dados

Muitos dos trabalhos existentes na literatura sobre visualização de mobilidade humana recorrem a coleções de leituras obtidas a partir de recetores GPS. Estes registos incluem dados além da própria posição, como a informação temporal do momento em que essa posição foi adquirida (o *timestamp*), a velocidade, e a orientação (*bearing*) entre outros atributos. Uma vez que este tipo de dados é frequentemente usado no estudo da mobilidade humana e como são dados fáceis de obter (de forma direta a partir de um recetor GPS), a nossa estrutura de conceitos deverá então adaptar-se a esta tipologia de dados de mobilidade. Por outro lado, a utilização de dados de GPS para representar a mobilidade humana origina outras preocupações. A primeira delas é a cobertura temporal dos dados, uma vez que é extremamente difícil recolher dados GPS em ambientes interiores e a maioria dos seres humanos passa grande parte do seu tempo nestes ambientes (Klepeis *et al.*, 2001). Assim, a maioria do tempo de observação não seria assim registada. O segundo problema é a dificuldade técnica em recolher dados de GPS em situações reais do nosso dia-a-dia. Uma das abordagens passa por atribuir aos indivíduos recetores de GPS (Van der Spek *et al.*, 2009) durante um período de tempo específico em que dura a experiência, mas que poderá provocar alterações nas rotinas diárias dos indivíduos. Esta abordagem tem a limitação do número de indivíduos observados não ser muito considerável e o tempo de observação ter de ser limitado. Como alternativa poderemos utilizar os dispositivos móveis, uma vez que atualmente os smartphones integram múltiplos sensores e estes sensores podem ser usados para observar os comportamentos dos indivíduos num referencial espaçotemporal.

Na implementação realizada, baseada na abordagem dos dados recolhidos por smartphones, pretendeu-se ultrapassar as limitações da recolha de dados baseada apenas em GPS. Foi então pedido

a um grupo de cinco utilizadores que instalassem uma pequena aplicação (*MOVE* ³⁵) nos seus smartphones *Android* permitindo recolher dados durante longos períodos de tempo e a partir de múltiplos sensores: coordenadas de GPS, os pontos de acesso de redes Wi-Fi nas proximidades e as células GSM às quais o smartphone se encontra associado. Uma das vantagens desta abordagem é não alterar os comportamentos habituais dos indivíduos, uma vez que não necessitam de prestar atenção à aplicação pois esta encontra-se em funcionamento (em *background*) sem necessidade de intervenção do utilizador. A segunda vantagem é a possibilidade de recolhermos dados tanto de ambientes exteriores (recorrendo aos três tipos de sensores) como interiores (apenas Wi-Fi e GSM).

De forma a existir um termo de comparação dos dados recolhidos com a realidade do dia-a-dia de cada utilizador, durante os dias em que se realizaram as recolhas de dados, os utilizadores foram preenchendo um diário com a descrição do instante de entrada e saída de cada um dos locais visitados (com o objetivo de ser possível ter uma validação dos resultados dos processos de transformação e identificação das situações anómalas na recolha). Com base num desses diários foi possível ter uma indicação do comportamento de mobilidade de um utilizador em particular (Figura 5.1), onde fica explícito que na maioria do tempo deste utilizador (cerca de 94%), este se encontra dentro de edifícios.

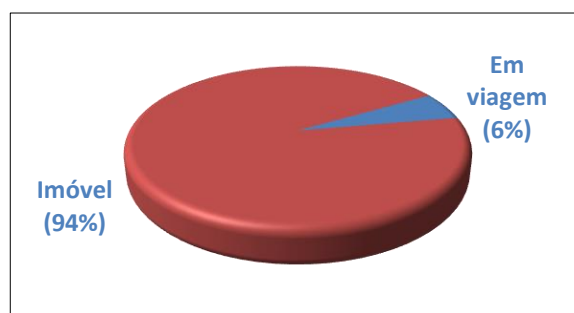


Figura 5.1 – Comparativo entre o tempo Imóvel e em Viagem

Os diários resumem assim os locais visitados e as viagens realizadas pelo utilizador durante o período de recolha de dados. A título de exemplo, apenas para contextualizar o perfil de mobilidade de um dos utilizadores, verifica-se que num período temporal de um mês, este visitou 22 locais distintos, realizando um total de 127 viagens entre esses mesmos locais. Com base ainda no diário foi possível

³⁵ Aplicação desenvolvida pela Universidade de Coimbra

identificar os cinco locais mais relevantes para este utilizador (onde a estadia foi maior tendo em conta o tempo acumulado ao longo do mês). A Figura 5.2 resume o tempo de estadia nos 5 locais mais relevantes quando comparado com os restantes 17 locais também visitados nesse mesmo mês.

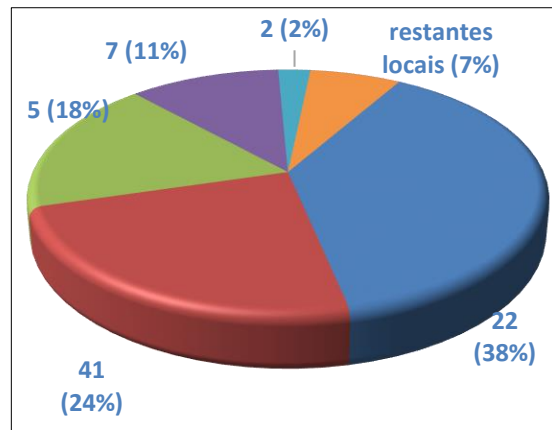


Figura 5.2 – Percentagem de tempo nos 5 locais mais relevantes e restantes locais

A imagem anterior indica o identificador de cada um dos locais mais relevantes e percentagem de tempo passado nesse local.

Na secção 2.2 verificámos que a mobilidade humana é um fenómeno repetitivo no sentido que visitamos com regularidade alguns locais e que existe um conjunto de locais onde passamos grande parte do nosso dia-a-dia. Embora não seja objeto deste trabalho o estudo e compreensão/explicação da mobilidade humana em si, como exemplo e usando os dados anteriores do diário, verificamos que estão de acordo com os resultados obtidos por González (González, Hidalgo e Barabási, 2008), 93% do tempo deste utilizador foi passado em apenas 5 dos 22 locais visitados.

Os dados recolhidos pela aplicação instalada nos smartphones refletem a mobilidade dos indivíduos nas suas normais atividades diárias. Realizarmos a recolha de dados de diferentes sensores para os mesmos instantes temporais visa, além da validação da estrutura de conceitos, validar a hipótese de a fusão de dados de diferentes sensores ser uma mais-valia para a construção de uma vista de mobilidade o mais completa possível quando comparada com a utilização única de apenas uma tipologia de dados.

As tabelas seguintes apresentam exemplos dos dados que foram recolhidos para um utilizador em particular, de acordo com a respetiva tipologia do sensor.

Timestamp	Latitude	Longitude	Altitude	Precisão	Velocidade
1338505272359	41.44869614	-8.29673100	295.600	12	0
1338505273495	41.44869614	-8.29673100	295.600	8	0
1338505274332	41.44869614	-8.29673100	295.600	8	0
1338754377428	40.75639343	-8.53647900	129.100	4	16.1

Tabela 5.1 – Exemplo dos dados GPS recolhidos

Os dados do sensor GPS incluem o *timestamp* em que foi registada a observação, as coordenadas latitude e longitude, a altitude a que se encontrava o artefacto naquele instante, a precisão da estimativa de cálculo das coordenadas e, por último, a velocidade instantânea.

Timestamp	RSSI	SSID	BSSID	Extra
1339151022344	-81	BUFFALO	00:16:01:a2:1b:74	b0:48:7a:c8:a8:f0,SaoPaio-TP,-88,2412
1339151027413	-86	BUFFALO	00:16:01:a2:1b:74	b0:48:7a:c8:a8:f0,SaoPaio-TP,-88,2412
1339151057826	-95	BUFFALO	00:16:01:a2:1b:74	b0:48:7a:c8:a8:f0,SaoPaio-TP,-88,2412
1339151226739	-94	BUFFALO	00:16:01:a2:1b:74	b0:48:7a:c8:a8:f0,SaoPaio-TP,-88,2412

Tabela 5.2 – Exemplo dos dados Wi-Fi recolhidos

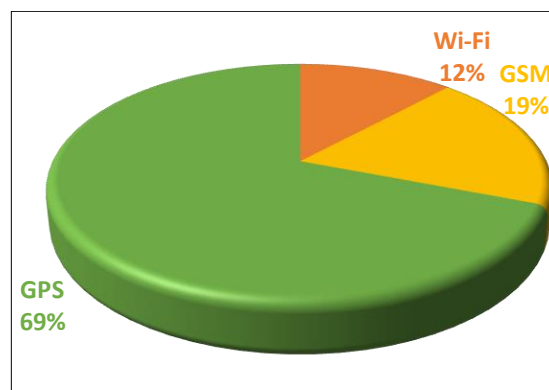
Relativamente aos dados Wi-Fi obtidos diretamente do sensor, mais uma vez contém o instante da aquisição, o RSSI (indicador da força do sinal recebido), o SSID que identifica a rede à qual o artefacto está no momento conectado, o BSSID que indica o endereço MAC do ponto de acesso ao qual o artefacto se encontra ligado e por último um campo denominado de Extra que guarda a informação dos pontos de acesso nas proximidades. Como uma rede Wi-Fi pode ser composta por múltiplos pontos de acesso é conveniente reter nos dados não só a identificação da rede (SSID) como o ponto de acesso (BSSID). A manutenção destes dois atributos são uma mais-valia para a inferência de por exemplo, zonas ou áreas dentro de um edifício, considerando que para o edifício em causa são necessários vários pontos de acesso de modo a uma total cobertura da rede Wi-Fi.

Os dados GSM que foram recolhidos pela aplicação eram compostos pelos seguintes atributos: o *timestamp*, mais uma vez para registar o momento temporal em que a observação se verificou; o CID que corresponde à identificação da célula GSM à qual o equipamento está associado; o LAC que identifica o código da área onde a célula se encontra instalada; o RSS que indica a potência do sinal recebido no equipamento. Além destes, foram ainda recolhidos outros atributos que, no entanto, por não terem dados ou por não serem relevantes para a representação da mobilidade, não foram por isso considerados para o efeito.

Timestamp	LAC	CID	RSS
1338505407902	8480	7944	21
1338505416897	8480	7944	20
1338506602417	8480	7942	-1
1338507557742	8480	7942	19

Tabela 5.3 – Exemplo dos dados GSM recolhidos

Como os vários *raw data* recolhidos, provenientes de diferentes utilizadores e diferentes períodos, encontravam-se em vários ficheiros, e de forma a facilitar o seu processamento, foi realizado um processamento intermédio (embora o mesmo não esteja descrito na estrutura de conceitos) que visou o carregamento destes ficheiros para uma base de dados MySQL, composta por três tabelas, uma por cada tipologia de dados. Como se trata de um processo simples de carregamento de dados, este processo não é considerado como um processo de transformação relevante (uma vez que não existe qualquer transformação ou adaptação dos dados) e como tal não é tido em consideração no desenho e definição da estrutura de conceitos.

**Figura 5.3 – Distribuição dos dados de acordo com a tipologia**

Uma dificuldade no processamento e arquivo de dados espaçotemporais de mobilidade prende-se com a enorme quantidade de dados que são recolhidos e necessários para a caracterização, mesmo que individual, da mobilidade humana. A título de exemplo, com a aplicação *MOVE* uma recolha de quinze dias de apenas um utilizador produziu 194.997 registos, tendo sido usados apenas três sensores. A Figura 5.3 ilustra a distribuição dos dados de acordo com a tipologia.

Embora a Figura 5.1 indique que o utilizador em análise passe a grande maioria do seu tempo dentro de edifícios, no entanto, devido principalmente à frequência de amostragem de cada um dos sensores, a Figura 5.3 poderia sugerir precisamente o contrário. Os dados associados à suspensão do movimento são assim em menor número quando comparados com os dados de GPS, associados ao

movimento. De facto, com uma frequência de amostragem de cerca um segundo para o sensor GPS, potencia um maior peso nos dados por parte deste sensor relativamente aos dois restantes. Mas ao analisarmos com mais atenção, pelo menos os dados relativos ao Wi-Fi, verifica-se que houve períodos confirmados pelo diário, de presença dentro de um edifício com pontos de acesso conhecidos e permitidos, que não resultaram em qualquer registo na recolha de dados. Posteriormente verificou-se que este fenómeno ocorria devido às configurações por defeito das ligações *wireless* dos dispositivos com sistema operativo *Android*. Sempre que o dispositivo verificava que as redes *wireless* não estavam a ser usadas para comunicação estas eram desativadas pelo sistema de forma a otimizar a bateria. Relativamente à recolha de dados GSM era de esperar também um número maior de registos e mais uma vez verificámos longos períodos em que não existe informação GSM. Neste caso não foi possível determinar o motivo do não registo, uma vez que o dispositivo estaria continuamente com cobertura e associado a uma célula GSM. Estes casos ilustram bem a realidade da observação da mobilidade humana no sentido em que é comum ocorrerem falhas e erros no processo de recolha de dados dos vários sensores.

5.2.2. Dos Dados em bruto às Observações

Com os dados carregados para uma base de dados específica, antes de iniciarmos o primeiro processo de transformação foi necessário criar dentro do MySQL uma nova base de dados exclusiva para a representação da estrutura de conceitos. A separação dos *raw data* dos restantes conceitos da referida estrutura visa manter os dados dos *raw data* intactos, sendo usados apenas para leitura durante o primeiro processo de transformação não voltando a ser necessária a sua utilização.

Assim o primeiro processo de transformação consiste no mapeamento dos dados de GPS, Wi-Fi e GSM num conjunto de *Observations*. Este processo é relativamente simples uma vez que consiste na leitura dos registos que estão disponíveis na base de dados dos *raw data* e na associação dos valores aos atributos de acordo com a tipologia dos dados. Isto é, partindo da definição apresentada no capítulo anterior, a estrutura de dados relativa ao conceito *Observation* seria a seguinte:

Id_Observation	Artefact	Timestamp	Location
-----------------------	-----------------	------------------	-----------------

Tabela 5.4 – Estrutura de dados do conceito *Observation*

sendo que o atributo *Location* poderia ser simbólico ou geométrico. Assim para que o processo de transformação possa contemplar de forma específica cada um dos tipos de dados, com atributos

distintos para cada um dos sensores (em particular relativamente ao atributo *Location*) a estrutura da tabela de dados implementada no MySQL e respetivos dados foi a seguinte:

Id_Observation	Artefact	Timestamp	Latitude	Longitude	BSSID	CID
1	289	1338508811	NULL	NULL	00:26:5b:2f:45:f8	NULL
23471	289	1338515429	NULL	NULL	NULL	7944
76060	289	1338810443	41.16278839	-8.63966084	NULL	NULL

Tabela 5.5 – Exemplo do resultado do primeiro processo de transformação

O atributo *Location* foi subdividido neste caso em quatro atributos (Latitude, Longitude, BSSID e CID) de forma a acomodar convenientemente os dados relativos ao posicionamento de acordo com cada um dos sensores utilizados. Esta adaptação em nada interfere com a definição do conceito e respetiva representação, e foi apenas realizada com o objetivo de facilitar e minimizar o tempo de acesso aos dados. Como esta tabela de dados é a mais volumosa em termos de quantidade de registos, a realização de consultas à base de dados com *queries* mais complexas iria originar um tempo de acesso muito elevado, dificultando assim o processamento da informação e a geração dos restantes processos de transformação.

Como em qualquer processo de aquisição ou recolha de dados, a existência de erros ou falhas na recolha ou mesmo no registo em ficheiro dos dados é uma situação habitual e com a qual teremos sempre que conviver. Neste trabalho, para não ser a exceção à regra, também existiram falhas na aquisição de dados originados pelas mais variadas razões: ou porque o smartphone ficou sem bateria e como tal não recolheu dados; ou recolheu dados, mas foram nulos, por exemplo, se recolheu dados de GPS em locais sem linha de vista com os satélites; ou a rede *wireless* foi desativada pelo sistema operativo (como vimos anteriormente), entre outras razões. Relativamente à falta de dados como seria de esperar não foi possível fazer nada em contrário. No entanto foi feita uma limpeza (ou um descartar de dados) no momento em que foi aplicado o primeiro processo de transformação. Se verificarmos que um registo apenas tem como válido o *timestamp* então não nos é útil para a análise.

5.2.3. Lugares e Estadias

A deteção automática de *places* que possam ser relevantes para um indivíduo (por exemplo, a sua casa ou o local de trabalho) ou para um grupo de indivíduos (por exemplo, um local popular numa localização urbana) é conhecida por *place learning*. Embora seja um conceito relativamente recente, na literatura é possível encontrar algumas implementações. No entanto, na sua maioria, estas

implementações recorrem apenas a um único tipo de observação de cada vez, por exemplo, ou GPS ou GSM. Mais recentemente o *place learning* tem vindo a ser aplicado em integrações de observações de múltiplos sensores (Chon e Cha, 2011).

Na implementação que foi realizada neste trabalho também houve necessidade de desenvolver um algoritmo de *place learning* que permitisse a utilização de observações obtidas dos três sensores: GPS, Wi-Fi e GSM. Este algoritmo, baseado num modelo probabilístico, inova na vertente de processamento simultâneo dos três tipos de observação, realizando desta forma uma fusão de dados enquanto agrega as observações de forma a identificar um *Place*. Esta abordagem é uma alternativa aos algoritmos de *clustering*, como por exemplo o DBSCAN (Ester *et al.*, 1996) ou SNN (Ert e Steinbach, 2003). Estes algoritmos definem a similaridade entre dois pontos multidimensionais utilizando funções de distância. Para conjuntos de dados onde cada um dos pontos sejam descritos por um par de coordenadas, a função de distância mais popular é a distância Euclidiana. No âmbito deste trabalho, enquanto as observações obtidas a partir do GPS incluem uma posição (par de coordenadas) o mesmo não acontece com as observações obtidas do sensor Wi-Fi e GSM que são posições simbólicas. Por este motivo era necessário definir uma função de distância diferente. A abordagem para a definição da função de distância teria que ter em consideração o conceito *Observation*: uma observação descreve o que um determinado sensor mediu num instante temporal e localização em particular, e não a localização em si. Considerando como exemplo uma observação baseada em GSM: a observação descreve que, num determinado instante, a célula GSM com maior potência de sinal nas proximidades é a célula C, e não que a observação foi obtida na localização da torre celular C. Como as células GSM têm uma cobertura alargada, a consequência é que duas observações baseadas em GSM mesmo obtidas em locais distantes podem mesmo assim ser semelhantes. No entanto, no processo de *place learning*, estas duas observações podem estar associadas a locais diferentes, o que significa que a célula C é “visível” a partir destes dois locais. De forma a considerarmos todas estas particularidades, o modelo escolhido para a função de distância não foi o Euclidiano, mas sim probabilístico. A probabilidade de dois pontos (observações O_i e O_j) pertencerem ao mesmo local:

$$P_{mesmolocal}(O_i, O_j) \quad (3)$$

é definida por um conjunto de 6 diferentes funções probabilísticas de forma a contemplar os três tipos diferentes de observações simultâneas:

Função Prob.	GPS	Wi-Fi	GSM
GPS	P_1	P_2	P_3
Wi-Fi	P_2	P_4	P_5
GSM	P_3	P_5	P_6

Tabela 5.6 – Matriz de probabilidades entre duas observações

Para a probabilidade P_1 , entre duas observações GPS, a distância Euclidiana pode ser um bom indicador. Se duas observações estão geometricamente próximas, então provavelmente referem-se ao mesmo local e, como tal, P_1 poderá ser descrito como:

$$P_1(o_i, o_j) = e^{-ED(o_i, o_j)/R_1} \quad (4)$$

onde $ED()$ é a distância Euclidiana entre as observações O_i e O_j , e R_1 é um parâmetro que relaciona a distância Euclidiana com a proximidade das duas observações. De salientar que P_1 toma o valor de 1 se as duas observações forem obtidas exactamente na mesma posição e tende para 0 se a distância Euclidiana tender para infinito.

Para as probabilidades P_2 e P_3 , a distância geométrica não poderia ser usada uma vez que tanto as observações Wi-Fi como GSM descrevem a localização de forma simbólica. O mesmo se aplica à probabilidade P_5 (Wi-Fi - GSM), uma vez que, embora de sensores diferentes, ambas representam de forma diferente a localização. Para contemplar estas três probabilidades foi considerada a diferença temporal entre as observações consecutivas de acordo com a seguinte hipótese: duas observações obtidas num curto intervalo de tempo podem referir-se ao mesmo local. As probabilidades P_2 , P_3 , P_5 são assim definidas da seguinte forma:

$$P_k(o_i, o_j) = e^{-|t_i - t_j|/R_k}, k = 2, 3, 5 \quad (5)$$

onde R_2 , R_3 e R_5 são parâmetros que relacionam a diferença temporal entre as observações e a proximidade entre as mesmas observações.

Para estimarmos a proximidade entre duas observações Wi-Fi é importante ter em consideração as características das redes Wi-Fi. A cobertura dos pontos de acesso das redes Wi-Fi é tipicamente reduzida e assume-se que tenha um raio de cobertura com cerca de 50 metros ou menos. Assim, se duas observações referirem a presença do mesmo ponto de acesso, poderemos assumir que foram obtidas no mesmo local. Se as duas observações referirem dois pontos de acesso diferentes, então deveremos ter em consideração a proximidade temporal como acontece por exemplo na probabilidade P_2 . Assim, para um par de observações Wi-Fi a probabilidade P_4 é definida como:

$$P_4(o_i, o_j) = \begin{cases} P_{sameAP}, & AP_i = AP_j \\ e^{-|t_i - t_j|/R_4}, & AP_i \neq AP_j \end{cases} \quad (6)$$

onde P_{sameAP} é a probabilidade de as duas observações referirem o mesmo local dado terem observado o mesmo ponto de acesso (AP). Esta probabilidade não foi propositadamente colocada com o valor de 1 uma vez que em locais com fraca cobertura da rede Wi-Fi o mesmo ponto de acesso poderá ser identificado em diferentes locais nas proximidades. Baixando ligeiramente este valor (foi utilizado o valor $P_{sameAP} = 0,95$), não se limita a dimensão do local à dimensão típica da cobertura Wi-Fi.

No caso de duas observações GSM, embora igualmente simbólicas como as observações Wi-Fi, não poderemos usar o mesmo modelo uma vez que as células GSM têm tipicamente uma área de cobertura muito maior do que a cobertura dos pontos de acesso Wi-Fi. Assim considerou-se apenas a proximidade temporal, no entanto, atribuindo relevância diferente caso as duas células identificadas sejam ou não diferentes:

$$P_6(o_i, o_j) = \begin{cases} P_{sameCell} \times e^{-|t_i - t_j|/R_6}, & cell_i = cell_j \\ P_{difCell} \times e^{-|t_i - t_j|/R_6}, & cell_i \neq cell_j \end{cases} \quad (7)$$

os parâmetros $P_{sameCell}$ e $P_{difCell}$ são usados como pesos na probabilidade, respetivamente com os seguintes valores: $P_{sameCell} = 1$ e $P_{difCell} = 0,8$.

Os valores atribuídos aos vários parâmetros das funções probabilísticas para a implementação que foi realizada foram os seguintes:

	GPS	Wi-Fi	GSM
GPS	$R_1 = 50$	$R_2 = 40$	$R_3 = 40$
Wi-Fi	$R_2 = 40$	$R_4 = 40$	$R_5 = 40$
GSM	$R_3 = 40$	$R_5 = 40$	$R_6 = 20$

Tabela 5.7 – Valores dos parâmetros utilizados nas probabilidades

Estes valores foram obtidos após vários testes realizados para um conjunto restrito de utilizadores e diferentes períodos de observação. É válido considerar que para outros utilizadores ou períodos mais alargados de análise seja necessário ajustar estes parâmetros.

Cada um dos *Places* extraídos com base neste modelo probabilístico poderá ter no máximo observações dos três diferentes tipos de sensores, com uma ou várias observações de cada tipologia.

A Figura 5.4 ilustra, sem nenhum rigor geométrico, o que poderá ser um potencial *place* composto por diversas observações de diferentes fontes de dados.

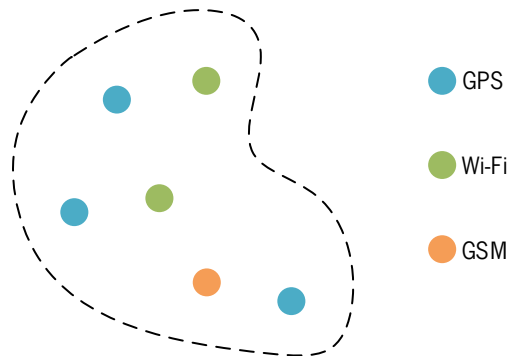


Figura 5.4 – Um *place* caracterizado por diferentes tipos de observações

A representação das observações Wi-Fi e GSM são meramente ilustrativas não correspondendo ao referencial geométrico, uma vez que as mesmas são inclusivamente simbólicas, ilustrando apenas que as mesmas caracterizam um determinado *place*. Assim um *place* é descrito por uma componente GPS, outra Wi-Fi e outra ainda GSM. A componente GPS é representada por um centróide de todas as observações GPS que foram adicionadas ao local e o *timestamp* da observação GPS mais recente que foi adicionada. A componente Wi-Fi é descrita por todos os BSSIDs das várias observações Wi-Fi que foram incluídas ao local e mais uma vez o *timestamp* mais recente de cada uma das *observations* correspondentes. A componente GSM tem uma representação similar à da componente Wi-Fi.

De acordo com a representação do conceito descrita na secção **Erro! A origem da referência não foi encontrada.**, um *Place* é então descrito por um conjunto de observações (cuja representação poderia ser descrita de acordo com a Tabela 5.8).

Id_Place	List of Observations
----------	----------------------

Tabela 5.8 – Estrutura de dados do conceito *Place*

No entanto, de forma a otimizar a codificação da implementação e considerando que temos três diferentes sensores com dados específicos, a implementação da representação do conceito *Place* foi feita da seguinte forma:

Id_Place	GPS		GSM		Wi-Fi		Centróide
	Coordenadas	Timestamp	CID	Timestamp	BSSID	Timestamp	Coordenadas

Tabela 5.9 – Estrutura implementada para o conceito *Place*

Cada um dos diferentes sensores tem assim uma lista de posições (geométricas ou simbólicas) e informação sobre o último *timestamp* da observação associada a esse sensor. Esta informação

individualizada sobre o último *timestamp* foi adicionada à representação de forma a facilitar o cálculo probabilístico baseado na distância temporal, quer seja entre sensores diferentes ou entre observações de igual tipologia. Foi ainda adicionada à informação base de um *place* o atributo centróide (obtido a partir do conjunto de observações GPS associadas a esse *place*) apenas com o objetivo de facilitar a representação geométrica do *place*, caso a mesma exista, uma vez que existirão situações em que a informação geométrica não existe.

Para a extração de *places* foi implementado um algoritmo em .NET, onde a cada nova observação é realizada uma comparação com as observações existentes associadas previamente aos locais já identificados (como representado na Figura 5.4). Se a probabilidade de uma observação ter sido recolhida num local já identificado for superior a um limite predefinido (P_{min}), então essa observação será considerada para ser potencialmente adicionada a esse local; caso contrário a observação dará origem a um novo local.

O algoritmo de extração de *places* pode ser descrito de forma sumária da seguinte forma:

Para cada nova observação
 Para cada place já extraído
 Calcular as várias probabilidades P_x
 Seleccionar a probabilidade mais elevada (P_{max})
 Se $P_{max} \geq P_{min} \rightarrow$ adiciona ao place correspondente
 Senão cria novo place
 Se a nova observação é GPS \rightarrow calcula centróide

Figura 5.5 – Algoritmo da extração do *Place*

Se existirem vários resultados das funções probabilísticas para a mesma observação cujo valor obtido seja superior ao P_{min} , nesse caso a observação é adicionada ao local cujo valor de P_{max} seja mais elevado.

A execução deste algoritmo a cada nova observação permite uma de duas hipóteses: enriquecer a caracterização de um *place* ou criar um novo *place*. Se a maior das probabilidades calculadas não for superior ao valor predefinido de P_{min} então teremos um novo *place*. No entanto se as observações forem recolhidas enquanto o individuo se encontra em movimento, é bastante provável para essa situação que o valor máximo das probabilidades não ultrapasse o P_{min} originando assim novos *places* com elevada frequência. A Figura 5.6 demonstra um exemplo desse cenário, onde numa zona geográfica com fraca probabilidade de ser identificado um *place* relevante foram extraídos a partir

das observações de GPS um número significativo de *places* candidatos (a zona geográfica em questão representa uma área de convergência entre as autoestradas A1 e A25 nas proximidades de Albergaria-a-Velha, não sendo por isso expectável a existência de paragens nestas vias).

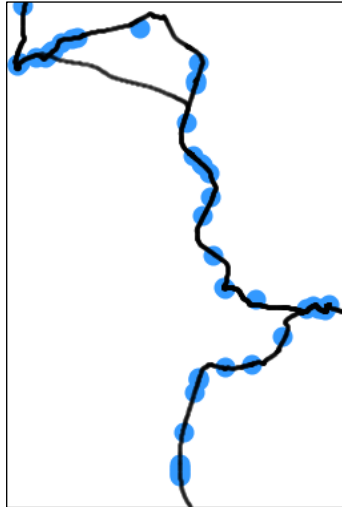


Figura 5.6 – *Places* candidatos de um indivíduo

A figura anterior (resultado de uma visualização produzida com recurso ao *Processing*) apresenta a azul os *places* candidatos extraídos a partir apenas das observações GPS. As mesmas observações permitiram reconstruir a preto os percursos realizados durante o período de análise para o mesmo utilizador. Mesmo ajustando os valores dos vários parâmetros usados no cálculo das probabilidades, em particular o R_1 , poderíamos diminuir em número os *places* mas teríamos sempre identificações que não correspondem a verdadeiros locais de estadia, mas sim a locais de passagem; além de que ao aumentarmos a área geométrica associada a cada *place* estaríamos a comprometer a distinção entre locais de relevo que se encontrem próximos. Por este motivo, na construção da estrutura de conceitos foi contemplada uma separação entre locais verdadeiramente importantes e os restantes, e adicionados à estrutura, os conceitos de *place* relevante e *place* candidato (Figura 4.8). Assim, até este momento o que o algoritmo produz são *places* candidatos que só posteriormente e mediante determinadas condições são elevados à categoria de *places* relevantes.

A métrica escolhida para a distinção entre *place* relevante e *place* candidato foi baseada no tempo de estadia nesse mesmo *place*. Considera-se como relevante para um indivíduo os locais em que os indivíduos passam grande parte do seu tempo, por exemplo os locais de trabalho, as nossas habitações, entre outros; são locais onde a estadia em termos temporais é elevada comparativamente a outros locais de visita rápida ou mesmo pequenas paragens não previstas ou tidas em consideração para o indivíduo, como é o caso das paragens no trânsito. Poderiam ter sido utilizadas outras métricas,

como por exemplo, considerar como relevante os locais com maior número de estadias, no entanto esta métrica pode levar a que sejam considerados como *places* relevantes as paragens recorrentes no trânsito e diminuir assim a importância dos locais que o indivíduo considera para si como verdadeiramente relevantes. Se considerarmos como exemplo um indivíduo que realiza frequentemente um trajeto com congestionamentos de tráfego recorrentes, basear a métrica no número de ocorrências de estadias, tornaria também esses pontos de congestionamento em locais relevantes. No entanto não significa que a métrica escolhida seja melhor ou pior, uma vez que a escolha deverá ter em consideração o tipo de análise que se pretende realizar, isto é, dependendo do tipo de análise existe uma noção diferente de relevância. Por exemplo, no caso das vistas individuais de mobilidade, o foco recai sobre os *places* relevantes para o indivíduo e nesse caso a métrica que foi usada nesta implementação poderá ser a mais indicada. No entanto, esta pode não ser a mais adequada se o cenário for a análise de vistas globais de mobilidade onde possa ser relevante para o planeador/decisor do espaço urbano, os locais onde existe uma forte presença de congestionamento de tráfego. Assim, é natural e ajustado que possam existir em simultâneo diferentes métricas em uso dependentes do tipo de análise em questão.

Como neste trabalho os dados disponíveis se referiam a mobilidades individuais e a amostragem não permite extrapolar conclusões globais, optou-se por implementar, nos processos de transformação, métricas mais direcionadas para as vistas individuais de mobilidade, salvaguardando sempre a possibilidade de poder ser usado outro tipo de métricas. As métricas em si não colocam em causa os conceitos nem os processos de transformação (principal preocupação neste trabalho), apenas poderão influenciar os resultados obtidos.

Tendo em consideração então a métrica selecionada, um *place* candidato apenas poderá passar a relevante se o tempo acumulado de estadia num local for igual ou superior a um valor predefinido (para efeitos de teste da implementação considerou-se que numa análise temporal de curta dimensão, dois minutos seria o valor mínimo a considerar). Este valor em si não é em termos temporais muito significativo, uma vez que num congestionamento de tráfego poderemos rapidamente atingir este valor. No entanto é um valor suficiente para se reduzir o número de *places* candidatos obtidos em parte pela mobilidade dos indivíduos.

Em resumo, e recorrendo novamente à Figura 4.8, o primeiro processo de transformação representado nessa figura tira partido do modelo probabilístico desenvolvido para agregar observações num conjunto alargado de *places* candidatos. Posteriormente, e de acordo com a definição do conceito

Place, existe um segundo processo de transformação que a partir dos *places* candidatos extrai as *stays* de cada *place*. Neste processo de transformação é determinado assim o tempo despendido em cada um dos *places* candidatos. Por último, um terceiro processo de transformação faz a distinção entre os *places* candidatos e os *places* relevantes com base, neste caso em particular, numa métrica temporal.

O algoritmo associado ao segundo processo de transformação, que permite a extração de *stays* pode ser descrito de forma sumária da seguinte forma (Figura 5.7), onde $P(O_i)$ e $T(O_i)$ representam respetivamente o *place* e o *timestamp* associado à *observation* O_i :

Para cada Place candidato(P)
 Se $P(O_i) = P(O_{i+1})$ e $(T(O_{i+1}) - T(O_i) < T_{max}) \rightarrow$ calcula Δt da *Stay* (S_j)
Para cada Stay extraída (S_j)
 Se Δt de $S_j < T_{min} \rightarrow$ elimina S_j
 Se $P(S_j) = P(S_{j+1})$ e $(T(S_{j+1}) - T(S_j) < T_{max}) \rightarrow$ agrega S_j e S_{j+1}

Figura 5.7 – Algoritmo da extração do *Stays*

No algoritmo deste processo de transformação assume-se que observamos um indivíduo no mesmo *place* (P) e que este não abandonou o local num intervalo de tempo entre observações consecutivas inferior a um T_{max} . Com base neste princípio é calculado então o tempo de presença nesse local e criada uma potencial *stay*. Desta forma não são considerados para efeitos de cálculo da *stay* dois tipos de situações: a primeira, em que existe mudança de *place* e como tal não poderia ser uma estadia; a segunda, quando as observações têm uma distancia temporal superior ao T_{max} dificultando afirmar com precisão que o indivíduo não se movimentou nesse período de tempo.

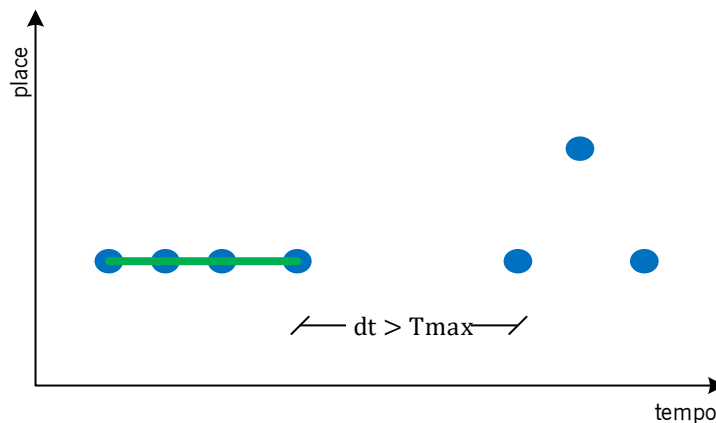


Figura 5.8 – *Stays* no espaço e tempo

Na Figura 5.8 temos uma ilustração das referidas situações. A linha a verde ilustra a *stay* extraída com quatro observações consecutivas no mesmo *place*, não podendo ser incluída a quinta observação em virtude do tempo entre observações ser superior ao T_{max} . As três últimas observações à direita na figura também não produzem nenhuma *stay* uma vez que existem variações do *place* embora sejam temporalmente próximas.

Numa segunda fase do algoritmo e tendo já sido extraídas as potenciais *stays*, é aplicado um filtro sobre esse conjunto de forma a eliminar estadias demasiado curtas (foi considerado um T_{min} de 10 segundos) que por serem de muito curta duração não poderão por isso ser consideradas verdadeiras estadias, mas sim breves paragens; e aplicado um mecanismo de agregação se entre duas potenciais *stays* consecutivas a distância temporal for inferior a T_{max} e o *place* for o mesmo.

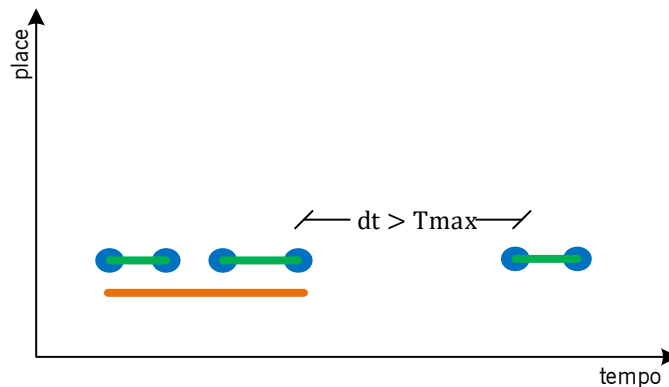


Figura 5.9 – Agregação de *stays* – exemplo 1

A Figura 5.9 demonstra a laranja a agregação feita entre duas potenciais *stays* que foram assim transformadas em apenas uma *stay*, uma vez que as mesmas são temporalmente próximas, o que já não acontece com a terceira *stay* que por distar mais do que o valor de T_{max} não pode assim ser agregada às duas restantes. Outra situação onde não é possível realizar a agregação de *stays* é o exemplo da Figura 5.10, que embora a distância temporal entre elas seja inferior ao valor de T_{max} , por terem *places* diferentes não poderão assim ser agregadas.

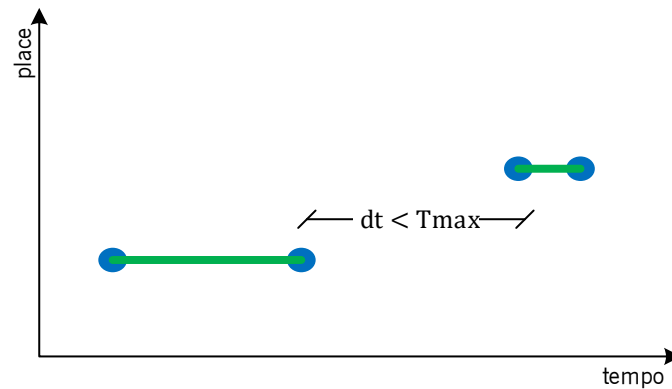


Figura 5.10 – Agregação de *stays* – exemplo 2

Como resultado da filtragem e agregação anterior temos então as *stays* finais extraídas com base nos *places* candidatos. A Tabela 5.10 apresenta o resultado da filtragem e agregação das *stays* consideradas como potenciais para um único utilizador durante um período de uma semana de observação.

Período de Análise	Potenciais Stays	Stays Finais
1 semana	485	210

Tabela 5.10 – Resultado da filtragem e agregação de *stays*

De acordo com a representação do conceito descrita na secção 4.1.4, uma *Stay* é descrita por um intervalo de tempo entre observações no mesmo local e a implementação desta representação foi feita da seguinte forma:

Id_Stay	Place	Timestamp_Initial	Timestamp_Final	Stay_Time
9357	1	1338516315	1338516339	24

Tabela 5.11 – Exemplo da estrutura de uma *stay*

Os quatro primeiros atributos derivam de forma directa da definição do conceito, instante inicial e final em que se observou um artefacto num determinado *place*, sendo adicionado para otimização de cálculo o atributo *Stay_Time* que regista em segundos a duração da estadia.

Por último, o terceiro processo de transformação associado à extração de *Places* e *Stays*, tem por base a identificação das *Stays* para permitir a diferenciação entre um *place* candidato e um *place* relevante.

O algoritmo associado a este terceiro processo de transformação, que permite a elevação de um *place* de candidato a relevante pode ser descrito de forma sumária da seguinte forma, onde $P(S_i)$

e $P(S_j)$ são os places associados a duas stays distintas e $\Delta t(P)$ o somatório de todas as stays do place P :

Para cada Place candidato(P)
Se $P(S_i) = P(S_j) \rightarrow$ actualiza $\Delta t(P)$
Se $\Delta t(P) \geq \Delta t_{min} \rightarrow P$ passa a relevante

Figura 5.11 – Algoritmo da elevação do *Place* de candidato a relevante

Assim, para cada um dos *places* candidatos identificados anteriormente são contabilizados todos os tempos do conjunto de *stays* existentes associados a esse *place*. Se o valor total calculado nesse somatório (Δt) for igual ou superior ao parâmetro predefinido Δt_{min} , então o *place* é considerado de relevante. Para efeitos desta implementação foi considerado um valor de Δt_{min} de dois minutos.

A Tabela 5.12 apresenta o resultado da aplicação dos três processos de transformação aplicados à extração de *Places* e de *Stays*.

Período de Análise	Places candidatos	Places relevantes
1 semana	3236	15

Tabela 5.12 – Resultado do terceiro processo de transformação

Como podemos facilmente constatar, para um curto período de análise de um único utilizador, o modelo probabilístico produziu uma quantidade significativa de *places* candidatos dos quais apenas 15 são efetivamente de importância para a construção da vista individual de mobilidade. Anteriormente tínhamos já constatado que a mobilidade do próprio utilizador é propícia à criação abundante de *places* candidatos (dado o elevado número de observações GPS) e como tal demonstrativa da necessidade dos dois últimos processos de transformação.

A Figura 5.12 apresenta geograficamente (apenas uma área parcial do mapa individual de mobilidade) o resultado da filtragem de *places* candidatos onde o tempo de estadia era reduzido, tendo em consideração este último processo de transformação.

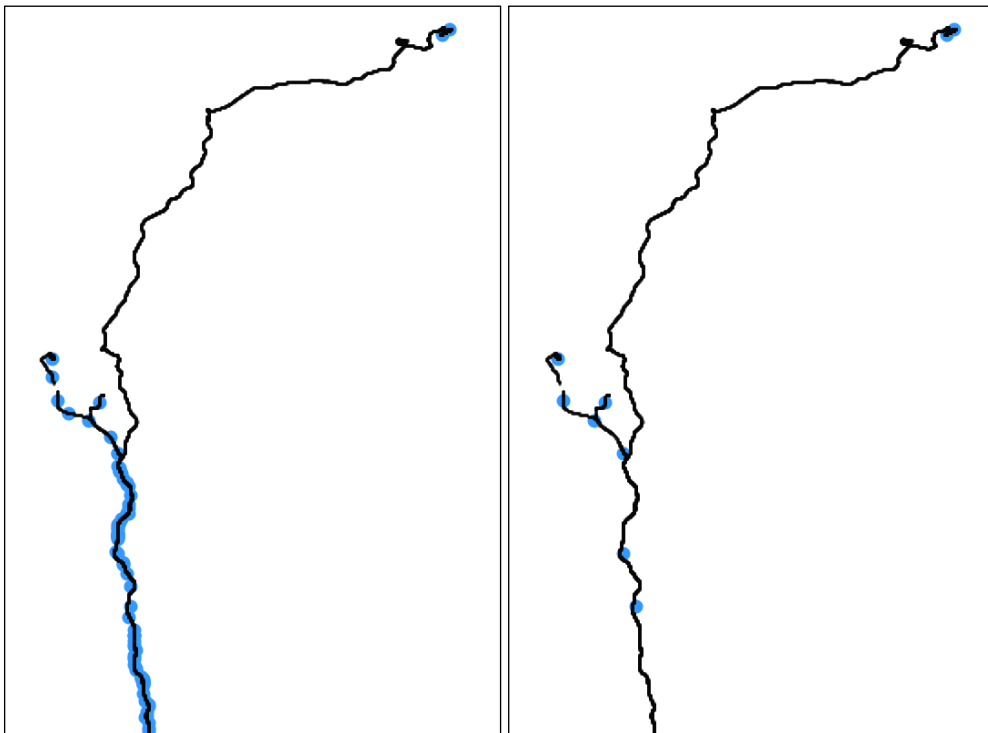


Figura 5.12 – *Places* candidatos (esquerda) e *places* relevantes (direita)

Comparando o resultado obtido pelo algoritmo de *place learning* com a informação registada no diário do utilizador verificam-se as seguintes situações. Primeiro, o diário do utilizador apontava para 10 *places* relevantes, no entanto o algoritmo detetou 15, dos quais apenas 6 têm correspondência temporal. Considera-se que os *places* têm correspondência temporal quando os valores temporais das *stays* (início e fim) se encontram dentro do intervalo temporal descrito no diário do utilizador. A correspondência não tem que ser perfeita, o início e o fim das estadias do algoritmo de *place learning* serem os mesmos que estão no diário, consideram-se para o efeito todos os casos em que total ou parcialmente o intervalo de tempo de uma *stay* se encontra contido no intervalo temporal apresentado no diário.

A Tabela 5.13 apresenta essa correspondência e permite verificar que existem locais reportados no diário que têm associado mais do que um *place*. Este resultado poderá indiciar que existem parâmetros no modelo probabilístico ou nos parâmetros associados à extração de *Stay* que são muito restritivos. No entanto se, por exemplo, aumentarmos a área de cobertura de um *place* (abrandando a restrição dos parâmetros) poderemos correr o risco de agregar num *place* único, observações que correspondem na realidade a diferentes *places* e que como tal se deseja a sua diferenciação. Era necessário um conjunto mais alargado de diários e dados recolhidos para com maior certeza se conseguir avaliar a origem destas ocorrências.

Place diário	Place extraído
1	1
11	1130;2381
2	162;137
25	2204
26	2210
27	-
28	-
29	-
30	2936
31	-

Tabela 5.13 – Comparativo entre os *Places* do diário e os *Places* extraídos pelo algoritmo

O facto de existirem dois ou mais *places* extraídos com associação a apenas um local real não é de todo preocupante, uma vez que o que ocorreu foi uma situação de *split* do local em dois ou mais *places*, mas mantendo sempre a associação correta. O que seria preocupante era ter um *place* extraído que estivesse associado a mais que um local real do diário (*merging*).

Diario_id	Nome	Início	Fim	Place extraído
1	Casa J – Guimarães	31-05-2012 21:45	01-06-2012 08:00	1
2	UM - Guimarães	01-06-2012 08:15	01-06-2012 18:10	162; 137
11	Casa C - Águeda	01-06-2012 19:55	03-06-2012 19:30	1130
25	Casa P - Gaia	03-06-2012 21:00	04-06-2012 07:00	2204
26	Fundação - Porto	04-06-2012 07:30	04-06-2012 16:30	2210
25	Casa P - Gaia	04-06-2012 17:00	04-06-2012 18:15	2204
27	Bar - Gaia	04-06-2012 18:30	04-06-2012 20:30	*
28	Gaia Shopping	04-06-2012 20:55	04-06-2012 21:10	*
25	Casa P - Gaia	04-06-2012 21:25	05-06-2012 06:55	2204
26	Fundação - Porto	05-06-2012 07:20	05-06-2012 16:15	2210
11	Casa C - Águeda	05-06-2012 17:40	05-06-2012 20:30	1130; 2381
29	Coimbra Shopping	06-06-2012 11:55	06-06-2012 12:30	-
30	Instituto Pedro Nunes	06-06-2012 12:40	06-06-2012 17:20	2936
31	Fórum Coimbra	06-06-2012 17:40	06-06-2012 17:52	-

Tabela 5.14 – Diário anotado com adição dos *Places* extraídos

A relação dos locais reais do diário com os *places* extraídos apresentada na tabela anterior é obtida a partir do resultado apresentado na Tabela 5.14, que apresenta o diário do utilizador para um

período de uma semana, ao qual foi adicionado o *place* extraído a partir do algoritmo de *place learning*, de acordo com a sobreposição do início e fim das *stays* com o início e fim anotado no diário.

Na Tabela 5.14 torna-se importante tentar identificar quais as potenciais razões para que o algoritmo de *place learning* não tenha conseguido extrair nenhum *place* relevante para alguns dos intervalos de tempo considerados no diário. Assim, não havendo *places* relevantes, procurou-se encontrar para esse período de tempo *stays* que confirmassem que existe pelo menos um *place* que tenha sido identificado como candidato. No caso dos locais identificados no diário como 27 e 28, estes não chegaram a ser identificados porque não existem observações para esse período temporal, muito provavelmente porque o dispositivo não se encontrava em funcionamento ou porque ocorreu alguma dificuldade na recolha de dados dos diferentes sensores. Durante os períodos em que durou a estadia nesses locais, por falta de observações, não foi assim possível a extração de *stays* e/ou *places*. Estes registos nos diários, para efeito de avaliação da *performance* do algoritmo de *place learning*, deveriam ser excluídos, uma vez que afetam de forma negativa os resultados sem que exista responsabilidade direta do algoritmo ou da sua implementação. Mais uma vez, como o conjunto de diários e dados recolhidos é muito reduzido, não permite em segurança realizar uma avaliação séria do desempenho do algoritmo de *place learning*.

Relativamente aos locais identificados como 29 e 31, neste caso foram extraídos *places* candidatos, no entanto os vários *places* candidatos tinham estadias muito curtas e realizadas apenas uma vez, isto porque o modelo probabilístico não encontrou semelhanças significativas entre as observações recolhidas e por isso foi gerando inúmeros *places* candidatos. O aumento de *places* candidatos diminui o tempo das *stays* e como tal o somatório do tempo de estadia não permitiu que fossem extraídas *stays* nesses intervalos de tempo.

Analisando mais em detalhe o que foi observado entre a saída do local 30 do diário até ao final da estadia no local 31 temos as seguintes *stays* potenciais (potenciais pois não chegaram a ser consideradas efetivamente como *stays* devido a não atingirem o valor mínimo de Δt_{min}), apresentadas na Tabela 5.15.

Id_Stay	Place	Timestamp_Initial	Timestamp_Final	Tempo
6749	2943	1339003337	1339003349	12
6753	2945	1339004708	1339004720	12
6759	2947	1339005083	1339005101	18
6764	2950	1339005147	1339005173	26

Tabela 5.15 – Extracto de *stays* potenciais

Como se verifica, num intervalo de tempo aproximado de trinta minutos, foram gerados *places* entre 2943 e 2950, o que corresponde à extração de 8 novos *places*. Destes apenas os 4 apresentados na tabela tiveram um tempo de estadia superior a um T_{min} e de forma individual, uma vez que não existem mais estadias para aqueles locais, o tempo de estadia máximo dos 4 registos apresentados é de 26 segundos, ainda muito longe do valor definido para o parâmetro Δt_{min} , o que significa que estes 4 *places* serão sempre candidatos e nunca considerados de relevantes.

Até ao momento verificámos os *places* que estão relacionados com os registos efetuados no diário do utilizador (correspondente a 8 dos 15 *places* extraídos), mas existem ainda 7 *places* que foram considerados relevantes pelo algoritmo, mas não para o utilizador, ao ponto de não os registar no diário.

Como não existe forma de os comparar com o diário, seguiu-se uma abordagem diferente e representou-se geograficamente os 7 *places* em falta. Um deles por falta de observações GPS não foi possível representar e como tal foi impossível de contextualizar. A Figura 5.13 contextualiza um dos restantes *places* e, embora não sendo considerado pelo utilizador como relevante no seu diário, verificou-se após apresentação e confrontação com o utilizador que efetivamente, não sendo relevante, era um dos locais frequentes de paragem, correspondente à Área de Serviço da Antuã na A1.

Os cinco restantes *places* identificados pelo algoritmo, como não foi possível ao utilizador identificar como locais que tenha visitado, e por se localizarem em vias de comunicação como se constata pelo exemplo da Figura 5.14, são muito possivelmente locais de passagem que: ou por sinalização luminosa, congestionamentos de trânsito, ou outra anomalia; tenha obrigado o utilizador a permanecer imóvel por um período de tempo igual ou superior ao limite Δt_{min} e como tal ter assim sido extraída uma *stay* e correspondente *place* relevante.

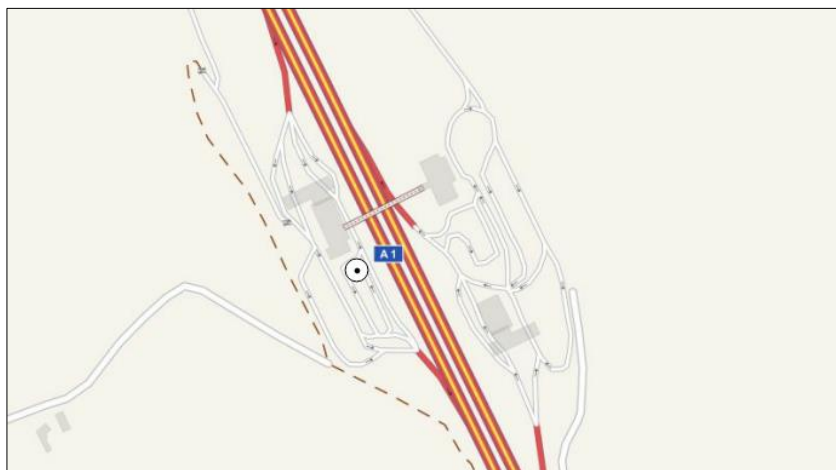


Figura 5.13 – Localização de um *place* extraído



Figura 5.14 – *Place* identificado numa via de comunicação

5.2.4. Suspensão do movimento

A definição do conceito *Time Leap* apresentada na secção 4.1.4 indica que o mesmo ocorre em situações onde a extração de *stays* não é aconselhada tendo em conta o limite definido pelo parâmetro T_{max} . A Figura 5.15 ilustra de forma gráfica este princípio.

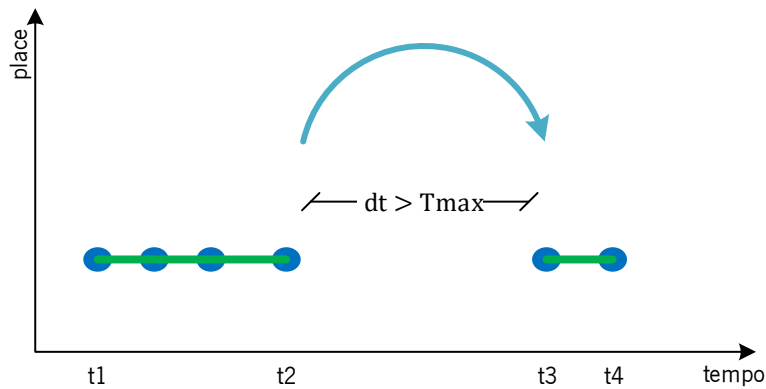


Figura 5.15 – Observações, *Stays* e *Time Leap*

As observações recolhidas entre os instantes temporais t_1 e t_2 permitiram extrair uma *stay* que ocorreu num *place* em particular. O mesmo acontece entre os instantes t_3 e t_4 , onde foi extraída uma nova *stay* para o mesmo *place*. No entanto as observações recolhidas no instante t_2 e t_3 aconteceram no mesmo *place* mas distam temporalmente entre si um período superior ao limite T_{max} pelo que não poderá ser considerada como uma possível *stay*. Esta distância temporal ocorre neste caso fruto da falta de observações entre os instantes t_2 e t_3 . Estamos assim, e de acordo com a definição do conceito, perante um *Time Leap*, uma vez que as duas *Observations* consecutivas no tempo distam um longo período temporal.

A implementação deste conceito, através de um processo de transformação é feita de forma quase direta, uma vez que perante duas observações consecutivas que referem o mesmo *place* se o intervalo de tempo entre elas for inferior ao T_{max} estamos perante uma *stay* e como tal foi processada anteriormente, senão estamos então perante um *time leap*. A Tabela 5.16 ilustra a estrutura usada na implementação do conceito assim como um exemplo da existência de um *time leap*.

Id_TimeLeap	Place	Timestamp_Initial	Timestamp_Final	Tempo
14	2210	1338893981	1338894242	261

Tabela 5.16 – Estrutura do conceito *Time Leap*

De salientar mais uma vez que, para o exemplo anterior, entre *Timestamp_Initial* e *Timestamp_Final* não existem observações que permitam afirmar se o artefacto se moveu ou não durante esse período temporal. Para um cenário de recolha de dados semelhante ao que foi realizado, bem como da implementação dos conceitos, a presença de *Time Leaps* representa uma anomalia que se deseja mínima. A forma automática e periódica com que a recolha de dados dos diferentes sensores foi realizada não deveria permitir, mesmo dentro de edifícios, que existam longos períodos de tempo sem a recolha de qualquer observação. É sabido que dentro de edifícios a receção de GPS é fraca ou mesmo nula, podendo existir períodos em que este sensor não tenha dados válidos, no entanto, mesmo não existindo uma rede Wi-Fi nas proximidades, seria de esperar a existência de observações GSM.

Assim, para este tipo de cenário aplicacional, a existência de *Time Leaps* é de maior importância para a caracterização dos dados que estão a ser analisados do que propriamente para efeitos de mobilidade dos artefactos. Com a implementação deste conceito foi assim possível verificar a falta de observações durante certos períodos de tempo, fruto de inúmeras situações possíveis, que para uma monitorização constante é um inconveniente significativo, mas não é de todo de estranhar. Desde a falta de bateria no *smartphone*, desativação das redes *wireless* para otimização do consumo da bateria, encerramento da aplicação devido a um erro, *smartphone* guardado num bolso ou pasta e como tal sem linha de vista com os satélites GPS, entre outras; são várias as potenciais fontes de falha na recolha ou inexistência de dados.

Para a caracterização da vista de mobilidade individual tendo em consideração quer a recolha de dados quer a implementação que foi realizada, direccionada principalmente para a validação da estrutura de conceitos, verifica-se que o conceito existe e pode ser implementado com recurso uma vez mais a um processo de transformação, mas os resultados obtidos não deverão ter, neste caso particular, a mesma relevância que os restantes conceitos na compreensão dos padrões individuais de

mobilidade. Poderá ser inadequado considerar que o artefacto se manteve num *place* tendo por base não as observações disponíveis, mas pelo contrário, a inexistência das mesmas. Inferir uma conclusão com base na ausência de informação poderá levar a conclusões incorretas ou precipitadas. No caso de se considerarem os *time leaps* na análise de mobilidade, em particular associada à ocupação de espaço, deverá ser sempre com a ressalva de que os mesmos são falsas *stays*. Nos vários trabalhos realizados por outros investigadores, o tratamento diferenciado das situações em que há falta de dados não é considerada e, frequentemente, assume-se que as pessoas se mantiveram no mesmo local durante esses períodos até que estejam disponíveis novas observações associadas a outros locais. É, por exemplo, o caso das análises efetuadas com base em registos de utilização das redes móveis celulares. Esta abordagem pode conduzir a resultados enviesados sobre a mobilidade humana.

5.2.5. Movimento

A implementação dos vários processos de transformação no algoritmo de *place learning*, desenvolvido com base no modelo probabilístico, permitiu a identificação de *places* relevantes bem como as *stays* nesses mesmos *places*. Estes dois últimos conceitos refletem a ausência de movimento. Por outro lado, para que ocorra uma mudança de local é obrigatório que os artefactos se movimentem entre locais. Assim a extração dos *places* relevantes é a base para a extração de outros conceitos da estrutura desenvolvida relativos à mobilidade, como por exemplo os *Elementary Movements* e os *Space Leaps*.

Vejamos a Figura 5.16, onde, para um dado utilizador, foram detetados vários *places* relevantes na cidade de Guimarães. Mesmo entre *places* relativamente próximos (à direita na figura, relativos à Universidade do Minho) existirão com certeza evidências de movimentações e como tal, observações dessas mesmas movimentações (no exemplo da figura as observações são representadas em tons de cinza).

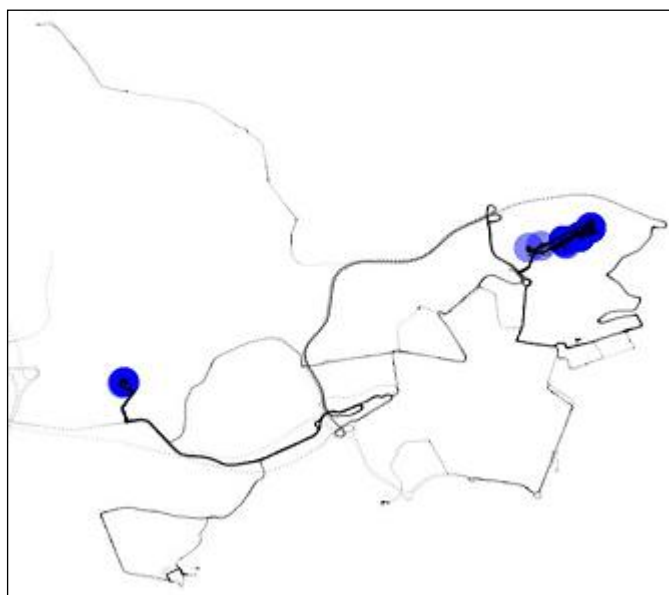


Figura 5.16 – *Places* relevantes

Assim, todas as observações que não estão diretamente associadas a *places* relevantes devem ser consideradas como potenciais evidências da mobilidade dos artefactos. Poderia ser feita uma análise a todas as observações para esse artefacto e aplicar processos de transformação para cada um dos conceitos associados ao movimento, no entanto, tendo em conta a grande quantidade de observações que são recolhidas para cada artefacto e uma vez que muitas destas já se encontram associadas a *places* relevantes não seria coerente considerar estas na extração de movimento quando as mesmas se referem a suspensão de movimento. A Figura 5.17 ilustra ao longo do tempo, de modo geral, a realidade associada a mobilidade de artefactos.

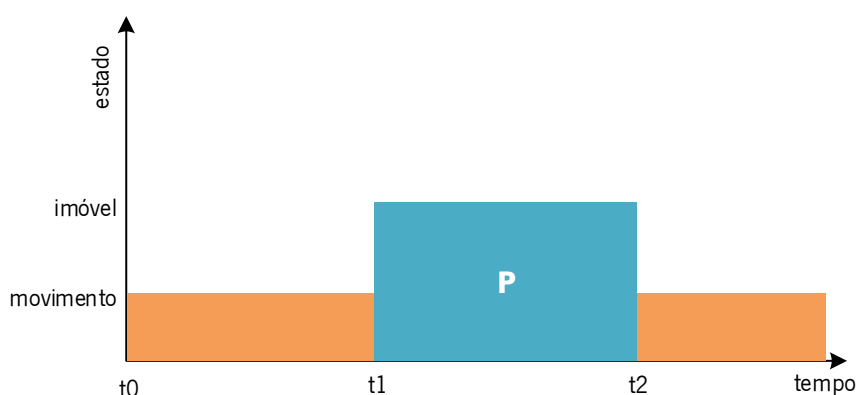


Figura 5.17 – Estados de mobilidade dos artefactos

Independentemente da natureza do artefacto, o mesmo poderá ser descrito num de dois estados de mobilidade: em movimento ou imóvel. Na figura, entre o instante t_0 e t_1 o artefacto

encontrava-se em movimento, ficando imóvel num *place* P entre t_1 e t_2 e retomando a sua marcha a partir do instante t_2 até voltar a imobilizar-se no mesmo ou noutra *place*. Tendo por base esta noção, as observações recolhidas entre os instantes t_1 e t_2 , por estarem associadas a um *place* e respetivamente ao estado imóvel, não acrescentam mais valia na extração dos conceitos associados ao movimento.

Assim o processo de transformação que permite a extração de *Elementary Movements* tem por base: as observações que não estão associadas a nenhum *place* relevante; estar em concordância com a descrição do conceito. Um *Elementary Movement* existe se, entre duas observações consecutivas no tempo, a diferença temporal não excede o limite T_{max} e se este par de observações têm *locations* diferentes. No caso da diferença temporal exceder o limite T_{max} , então não temos um *Elementary Movement* mas sim um *Space Leap*.

A implementação de ambos os conceitos é em tudo semelhante, uma vez que, o que os distingue é apenas o limite temporal entre observações. Assim um *Elementary Movement* pode ser representado da seguinte forma:

Id_Movement	Location_Start		Location_End		Timestamp_Initial	Timestamp_Final
	Latitude	Longitude	Latitude	Longitude		
5	41.4419585	-8.3212745	41.441921	-8.3212208	1322835034	1322835035
8	41.4419156	-8.3211565	41.441894	-8.3210974	1322835036	1322835037

Tabela 5.17 – Exemplo da representação dos *Elementary Movements*

Como facilmente constatamos no exemplo anterior, a *location* é geométrica. A definição do próprio conceito está muito associada às observações GPS. Representa pequenos movimentos entre *locations*, podendo-se facilmente interpolar posições intermédias. Dos vários sensores cujos dados foram recolhidos, os dados de GPS são aqueles que melhor representam esse fenómeno. Relativamente às observações GSM e Wi-Fi, como fornecem *locations* que são simbólicos, não permitem a interpolação de posições intermédias, além de que o intervalo de amostragem para estes dois sensores é significativamente elevado quando comparado com a ideia subjacente ao conceito de *Elementary Movement*. No entanto, se a frequência de amostragem destes sensores for igual ou inferior à distância temporal entre observações consecutivas, distância esta definida pelo modelo probabilístico, é perfeitamente natural que possam existir *Elementary Movement* com base em observações simbólicas.

Relativamente ao *Space Leap* e respetiva representação, este pode ser representado de forma em tudo semelhante ao *Elementary Movement*:

Id_SpaceLeap	Location_Start		Location_End		Timestamp_ Initial	Timestamp_ Final
	Latitude	Longitude	Latitude	Longitude		
404	41.4472746	-8.2990872	41.451067	-8.2931864	1322835532	1322835710
554	41.4528107	-8.2890933	41.452762	-8.2890665	1322838709	1322838945

Tabela 5.18 – Exemplo da representação dos *Space Leaps*

Com a informação disponível associada ao conceito *Elementary Movement* facilmente se conseguem obter representações idênticas à da Figura 5.16, recorrendo à representação por exemplo de apenas o atributo *Location_Start*, ou se preferirmos uma representação mais exata do conceito, recorrer aos atributos *Location_Start* e *Location_End* para produzir uma visualização baseada em pares de *locations* consecutivas como apresentado na Figura 5.18.



Figura 5.18 – Exemplo de *Elementary Movements*

A Figura 5.18 apresenta a verde os vários *Elementary Movement*, onde a seta indica o sentido do movimento entre diferentes *locations*. Com o objetivo de melhor ilustrar as movimentações, a Figura 5.19 apresenta a mesma informação, no entanto com maior detalhe.

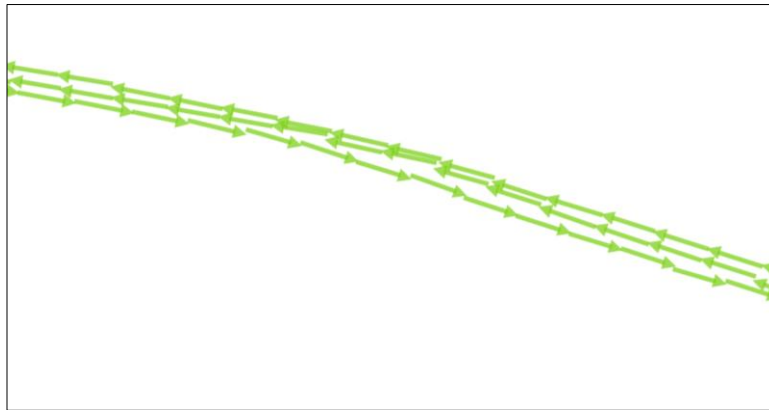


Figura 5.19 – Detalhe com vários *Elementary Movements*

A principal vantagem da representação da movimentação entre *locations* comparativamente com a simples representação da localização do artefacto é a possibilidade de ilustrar a direção do movimento, podendo ser relevante para a distinção entre que vias de tráfego (como se verifica na Figura 5.19) têm, por exemplo, maior congestionamento.

Na Figura 5.18 além de ser possível a reconstrução das vias de tráfego com base nos *Elementary Movements* do utilizador é ainda possível identificar as rotas mais usadas pelo utilizador em questão através da densidade espacial de *Elementary Movements*. De forma a realçar quais as rotas mais comuns do utilizador, foi usada uma opacidade reduzida nas configurações da visualização, permitindo assim que através da sobreposição das rotas a opacidade aumentasse diretamente de acordo com o número de ocorrências. Desta forma torna-se mais evidente, para quem interpreta a informação representada, quais os padrões de mobilidade mais comuns para o utilizador em questão.

Independentemente da opção de representação da informação, a estrutura de dados associada ao conceito permite múltiplas formas de visualização, uma vez que contém os dados mínimos e necessários para as necessidades representacionais.

Contrariamente aos *Elementary Movements* em que é possível seguir os movimentos realizados pelo utilizador ao longo do tempo, no caso dos *Space Leaps* a representação ilustra “saltos” maiores ou menores no espaço, em resultado da diferença temporal entre as observações consecutivas ser superior ao T_{max} . A Figura 5.20 ilustra exemplos de *Space Leaps* com base nos dados recolhidos para um utilizador em particular, onde os arcos a verde representam esses movimentos e a cinza as *locations* de origem e destino dos movimentos.

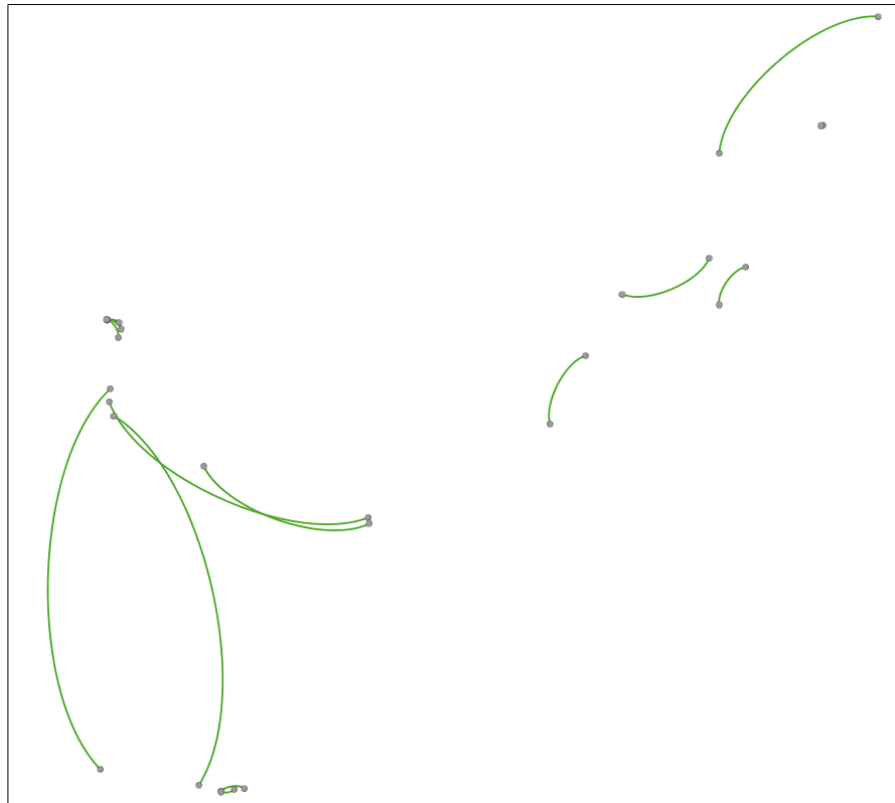


Figura 5.20 – Exemplos de *Space Leaps*

Como acontecia com os *Elementary Movements*, também na representação gráfica dos *Space Leaps* era desejável ter as setas com a indicação do sentido do movimento, quer para facilitar a leitura, quer para ir ao encontro da definição do conceito. No entanto, por não ser possível na ferramenta *Gephi* atribuir a direção a arcos, mas apenas a segmentos de reta, não foi possível incluir o sentido do movimento na representação da Figura 5.20.

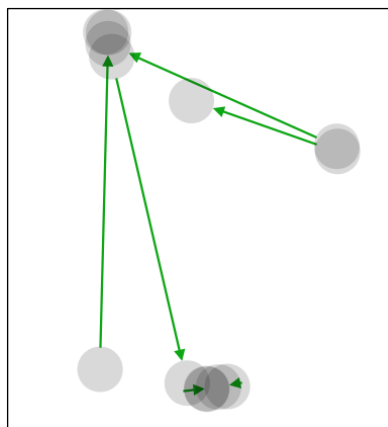


Figura 5.21 – Exemplo da dificuldade na representação no *Gephi*

A Figura 5.21 ilustra como seriam os *Space Leaps* com o sentido do movimento, mas sem a representação em arco. Como não é possível ter as duas desejadas características para a representação gráfica dos *Space Leaps*, optou-se pela representação em arco de forma a ser mais fácil a distinção entre *Space Leaps* e *Elementary Movements*.

Assim como acontecia com o conceito *Time Leap*, no caso dos *Space Leaps* os mesmos também ocorrem por não existirem observações em maior número que permitam extrair *Elementary Movements* em vez de *Space Leaps*. Mas para efeitos de análise de mobilidade, um *Space Leap* tem relevância pois demonstra que o artefacto se movimentou (embora não sendo possível dizer com exatidão o caminho realizado, nem se realizou uma *stay* entretanto), contrariamente aos *Time Leaps* que seriam *loops* sobre o mesmo *place*. De realçar que estas considerações se aplicam a este cenário particular de implementação, podendo haver ligeiras alterações em outros cenários aplicacionais ou com outras tipologias de dados.

A Figura 5.20 representa *Space Leaps* com base em observações geométricas, relevantes para a compreensão geográfica da mobilidade quer individual, quer global. Relativamente às observações simbólicas, quando não associadas a um *place* com informação geométrica, torna-se mais difícil a contextualização do *place*. No entanto podem permitir extrair outra informação igualmente válida para a caracterização dos mapas individuais e globais de mobilidade. Por este motivo foi experimentado representar *Space Leaps* apenas com as observações Wi-Fi. De forma direta dificilmente poderão acrescentar informação importante relativamente à mobilidade num espaço urbano, mas permite ilustrar a maior ou menor atração entre posições que podem ser do mesmo edifício ou de todo o espaço urbano visitado pelo artefacto.

A Figura 5.22 ilustra através de grafos não direccionados a forte ou fraca interação entre *places* apenas caracterizados por observações Wi-Fi. Mais uma vez, nesta figura, a opacidade dos arcos associados aos *Space Leaps* ilustra as ocorrências de transição entre locais. Com base nesta figura é perceptível quais transições entre *places* são mais relevantes em termos de número de ocorrências quando comparadas com as restantes. Esta informação permite caracterizar entre que locais o artefacto se movimenta mais. Se com base nas características do *place* (identificação dos pontos de acesso das observações que o definem, períodos temporais em que ocorreram as *stays*, entre outras) for possível associar o *place* a um local de relevo para o artefacto poderemos determinar se os *places* relevantes que foram identificados correspondem efetivamente a locais de importância para o artefacto (casa, trabalho, entre outras).

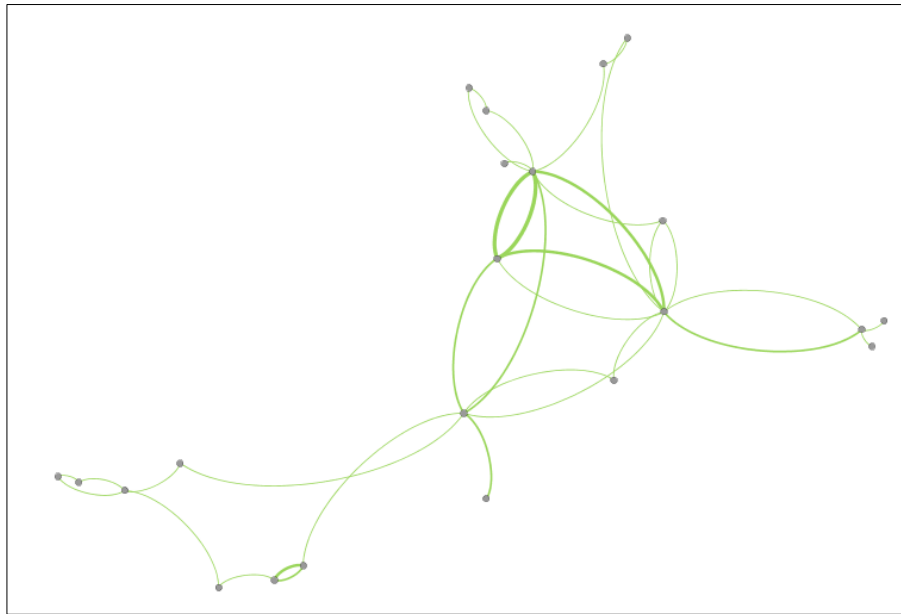


Figura 5.22 – *Space Leaps* apenas simbólicos

Assim, mesmo tendo unicamente observações simbólicas, originando *places* sem componente geométrica, é ainda possível recolher informação de mobilidade que no mínimo ajude à compreensão ou caracterização da mobilidade individual. A dificuldade está apenas na contextualização dos *places* no aspecto geográfico do espaço urbano, informação relevante quando se pretende analisar o espaço urbano como um todo.

5.2.6. Trajetória

Por último, temos o processo de transformação associado ao conceito *Trajectory*. Também este processo é relativamente simples, uma vez que de acordo com a definição do conceito é uma sequência ordenada temporalmente dos vários *Elementary Movements* e *Space Leaps* que ocorreram ao longo do intervalo de análise.

A Figura 5.23 ilustra um pequeno exemplo de uma *Trajectory* obtida através da representação dos dois conceitos associados ao movimento. As pequenas setas a azul e os nós a cinza ilustram as origens e destinos dos *Elementary Movements* e a vermelho estão representados os vários *Space Leaps*³⁶.

³⁶ Uma vez mais por limitação do *Gephi*, não é possível adicionar o sentido do movimento aos *Space Leaps*

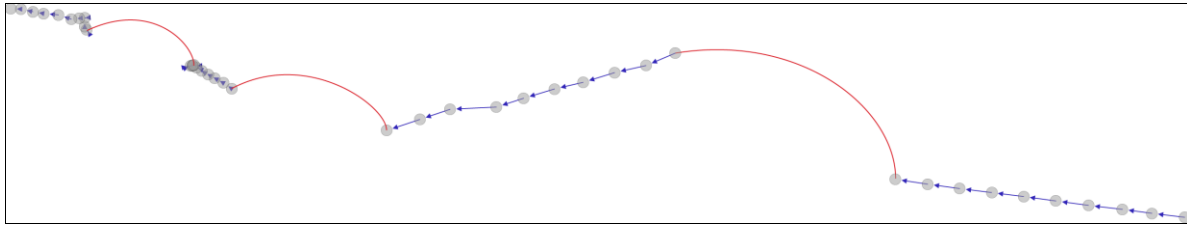


Figura 5.23 – Exemplo de *Trajectory*

A figura anterior ilustra de forma resumida um pequeno movimento realizado por um artefacto num dia, onde é de fácil percepção o sentido do movimento, no entanto a representação poderá ser mais complexa se analisarmos a mobilidade mensal de um artefacto, como demonstra a Figura 5.24.

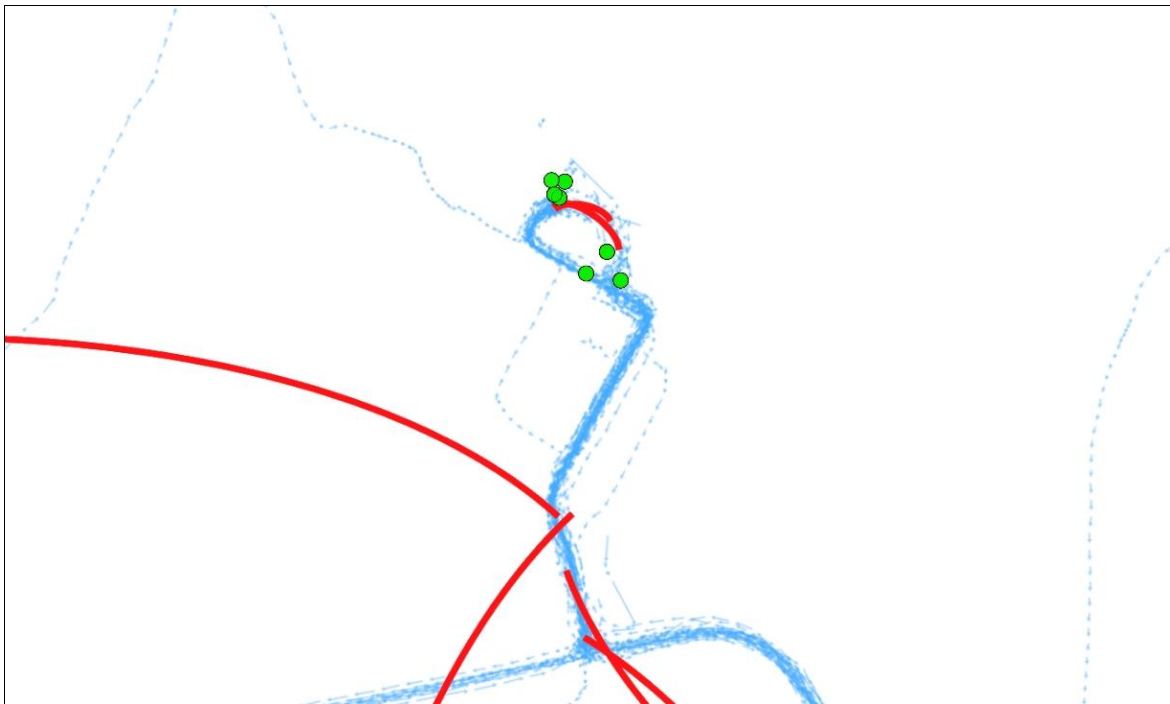


Figura 5.24 – *Places*, *Elementary Movements* e *Space Leaps*

Na figura estão representados, a azul, todos os *Elementary Movements* realizados durante um mês numa área geográfica restrita; a vermelho os *Space Leaps*; e, a verde, os *Places*. Estes últimos, não fazendo parte do conceito *Trajectory*, apenas foram aqui representados de forma a contextualizar os restantes conceitos. Torna-se assim mais legível a origem ou destino dos *Elementary Movements* e *Space Leaps*, e demonstra que estes movimentos não se realizam apenas entre *places* relevantes, mas sim entre *locations* que podem ou não estar associadas a *places*.

5.3. Redes Sociais

Na secção 5.2 foi apresentada a implementação da estrutura de conceitos, quer ao nível dos próprios conceitos, bem como ao nível dos vários processos de transformação necessários para a adequação da informação aos conceitos. O objetivo dessa implementação visava a própria evolução/validação da estrutura de conceitos e do modelo probabilístico associado à fusão de dados provenientes de múltiplos sensores. A experimentação realizada permitiu ainda demonstrar o potencial da estrutura de conceitos assim como dos resultados que se conseguem obter ao nível da análise de mobilidade humana de forma a produzirmos mapas individuais de mobilidade. No entanto o cenário aplicacional dessa implementação contempla a recolha de grandes conjuntos de dados obtidos pela monitorização constante de um artefacto ao nível da sua mobilidade.

Na literatura existem vários exemplos de estudos de mobilidade com recurso às redes sociais, estudos esses que retiram partido das publicações feitas em uma ou várias redes sociais e que, com base no conteúdo das publicações ou na informação extra associada às publicações, permitem inferir padrões de mobilidade, sobretudo globais, existentes num espaço urbano.

Mas contrariamente à monitorização contínua do artefacto, as publicações nas redes sociais são feitas de forma esparsa no tempo, sem um padrão específico. Como a natureza dos dados é neste caso muito diferente da implementação realizada na secção anterior, era primordial validar também se a estrutura de conceitos desenvolvida suporta dados recolhidos a partir das redes sociais e verificar até que ponto os resultados obtidos ao nível de análise de mobilidade permitem ou não extrapolar reais comportamentos existentes no espaço urbano.

A experimentação contou com dados recolhidos de duas redes sociais com o objetivo da mesma não ficar muito condicionada quer pela tipologia de dados de apenas uma rede social, quer pelo comportamento/motivação dos utilizadores relacionados com o ato em si de publicar nessa mesma rede social. Ao considerarmos duas redes sociais com características relativamente distintas pretendemos assim diversificar a experimentação e avaliar se existem diferenças na análise que uma e outra potenciam.

5.3.1. Aquisição dos Dados

De forma a realizar a validação da estrutura de conceitos para os *raw data* recolhidos das redes sociais foi necessário encontrar um conjunto de dados previamente recolhidos, uma vez que era

relevante verificar de que forma o uso de conjuntos de dados já utilizados previamente por outros investigadores permitiria obter os mesmos resultados ou resultados complementares.

Os dados inicialmente seleccionados foram disponibilizados por Cheng³⁷ (Cheng, Caverlee e Lee, 2010). Uma primeira análise destes dados, pelo facto de os mesmos terem sido usados por estes investigadores em anteriores publicações, indiciava que os mesmos dados pudessem vir a ser usados também para validação da estrutura de conceitos. Compostos por dois conjuntos de dados, de treino e de teste, onde para treino existiam 115 mil utilizadores da rede social *Twitter* e quase 4 milhões de *tweets* registados, e para teste um menor número de utilizadores, apenas 5 mil, mas um número de publicações superior, cerca de 5 milhões. Estaríamos perante dois *datasets* com alguma dimensão, resultante de uma recolha de quatro meses de publicações o que seria um bom ponto de partida ao nível dos *raw data*.

No entanto, após o carregamento dos dados para o conceito *Observation* verificou-se que os dados não seriam indicados para a validação que se pretendia. Embora os dados contivessem informação temporal e espacial, o que era um pré-requisito inicial, verificou-se que a informação temporal das publicações estava em condições de utilização, mas a componente espacial destes *datasets* referia-se ao posicionamento da conta do utilizador e não da publicação. Conclusão, as coordenadas GPS que existiam não se referiam ao local onde foi realizada a publicação. Portanto, todas as publicações de cada utilizador teriam sempre a mesma *Location*. Além da informação existente nada indiciar ao nível da mobilidade, ao avançar-se com a validação os resultados seriam deturpados e inúteis. Chegaríamos à conclusão que os utilizadores desta rede social nunca se movimentariam e apenas teríamos *Places* e *Time Leaps*. Com base nestes resultados foi necessário procurar *datasets* de publicações em redes sociais onde a informação espacial estivesse associada à publicação e não à conta do utilizador.

Como o *Twitter* alterou os termos e condições relativos ao uso e disponibilização de dados, em particular por terceiros³⁸, neste momento é necessário um maior esforço para se conseguir obter

³⁷ Disponível em: https://archive.org/details/twitter_cikm_2010 (visitado em Agosto de 2016)

³⁸ Termos e condições do *Twitter*: <https://dev.twitter.com/archive/terms/api-terms/diff-20130702> (visitado em Agosto de 2016)

datasets de publicações nesta rede social, principalmente porque se pretendem dados georreferenciados. Desta forma, muitos dos *datasets* usados em trabalhos de investigação que anteriormente estariam disponíveis, já não são possíveis de usar devido a restrições legais. Por este motivo, foi necessário recorrer a alternativas para a obtenção de um *dataset* com os requisitos necessários para a validação da estrutura de conceitos.

A alternativa encontrada foi aceder aos dados do *Twitter* através da Dandelion³⁹ API disponibilizada pela Spaziodati⁴⁰. Os dados recolhidos estão restritos à cidade de Milão, Itália, e devido a limitações de acesso livre aos dados, apenas foi possível recolher 100 mil publicações. Embora os dados estejam anonimizados, o processo de anonimização garantiu que o identificador associado a um utilizador seja o mesmo ao longo de todo o período de recolha.

Os *raw data* recolhidos estavam representados no formato JSON estruturado da seguinte forma:

```
{ "items": [
  { "language": "it", "created" : "2013-11-07T22:48:50",
    "geometry":{"type":"Point","coordinates":[longitude,latitude]},
    "timestamp":1383860930,
    "municipality": { "name" : "Milano", "acheneID" : "link"},
    "entities" : ["link", "link", "link"],
    "user" : "d96eca82bd"},
  ]
}
```

Figura 5.25 – Estrutura do objecto JSON

Para que os dados pudessem ser convenientemente trabalhados foi desenvolvido um módulo de *software* que fez o *parser* de toda a estrutura JSON para uma tabela de dados no MySQL. Dos dados existentes nos *raw data* apenas foram importados os atributos necessários para a análise de

³⁹ Disponível em: <https://dandelion.eu> (visitado em Agosto de 2016)

⁴⁰ Disponível em: <http://spaziodati.eu/en/> (visitado em Agosto de 2016)

mobilidade. A Tabela 5.19 apresenta a estrutura da tabela de dados criada para importar os *raw data* assim como alguns exemplos dos dados importados.

Timestamp	User	Latitude	Longitude
1383860930	d96eca82bd	45.49581987	9.1737967
1383860933	0c2d47b049	45.49319593	9.20732436
1383860938	36280c79d4	45.47882267	9.18660586
1383860942	ab925ae26f	45.45932174	9.21437243

Tabela 5.19 – *Raw data* importados

Com o objetivo de se verificar se com outras redes sociais as condições e posteriores conclusões são em tudo semelhantes, foi considerada a utilização também de dados da rede social *Foursquare*. Como estas duas redes sociais têm finalidades distintas é natural que os pressupostos de publicação nas mesmas também sejam diferentes, logo, os dados recolhidos serão nesse caso reveladores de comportamentos diferentes.

Neste caso o acesso aos dados foi feito de forma idêntica à primeira tentativa de recolha de dados do *Twitter*, reutilizar *datasets* de outros investigadores. Assim, com base nos dados usados e disponibilizados por Sarwat⁴¹ (Sarwat *et al.*, 2014), foram selecionados para a validação da estrutura de conceitos apenas dois *datasets* dos cinco que se encontravam disponíveis. A opção recaiu apenas sobre dados de *checkins* e de *venues*. No caso do *dataset checkins*, correspondente a um milhão de registos, os atributos disponíveis eram: *id*, que identifica o registo; *user_id*, com a identificação do utilizador que realizou o *checkin*; *venue_id*, que permite a identificação do local onde foi realizado o *checkin*; e o atributo *created_at* que indica o instante temporal do *checkin*. A Tabela 5.20 apresenta os atributos e exemplos de registos de dados do *raw data* que foram importados para uma base de dados auxiliar no MySQL.

⁴¹ Disponível em: https://archive.org/details/201309_foursquare_dataset_umn (visitado em Agosto de 2016)

id	user_id	venue_id	created_at
183647	1	51	2011-12-12 19:54:14
327105	1	84329	2011-12-25 00:47:33
609711	1	7489	2012-01-29 17:06:02
615245	1	64	2012-01-29 22:56:11

Tabela 5.20 – *Raw data checkins*

De salientar que relativamente ao *dataset checkins* existiam ainda mais dois atributos que não foram considerados por não serem relevantes para este processo. Estes dois atributos, *latitude* e *longitude*, embora incluídos no *dataset* de *checkins* não refletiam o posicionamento do *checkin* mas antes a localização associada à conta do utilizador, algo que tínhamos verificado previamente com outros *raw data* mas da rede social *Twitter*. Por esse motivo seriam dados falsos para a análise de mobilidade e como apenas serviriam para sobrecarregar a base de dados não foram por isso considerados.

Relativamente às *venues* ou locais onde se realizaram os *checkins*, o *dataset* é composto também por um milhão de registos com os seguintes atributos: *id*, que identifica a *venue*; *latitude* e *longitude* relativos ao local onde se realizaram os *checkins*. A Tabela 5.21 apresenta a estrutura dos atributos assim como exemplos de alguns dos registos que foram importados.

id	latitude	longitude
1	44.88201056	-93.21236372
2	44.883169	-93.213687
3	44.88345492	-93.21431637
4	44.88138718	-93.21380138

Tabela 5.21 – *Raw data venues*

5.3.2. Dos Dados em bruto às Observações

Realizada a importação dos dados *raw data* de duas redes sociais, o passo seguinte passou pela realização do primeiro processo de transformação para *Observations* como fizemos anteriormente, embora agora com os ajustes necessários de acordo com a fonte de dados.

Assim o primeiro processo de transformação consistiu no mapeamento dos dados do *Twitter* e do *Foursquare* num conjunto de *Observations*. Como os *raw data* importados têm identificadores de utilizadores de acordo com cada uma das redes sociais e existiria sempre a possibilidade de sobreposição dos identificadores associados aos artefactos, era necessário encontrar uma solução para

este problema. Foram consideradas duas soluções. A primeira seria ter duas estruturas semelhantes para a representação das observações, onde cada tabela corresponderia a uma rede social. A segunda solução, e talvez a mais recomendável, seria a criação de um atributo auxiliar que indicasse qual o sensor associado à observação. Esta abordagem tem a vantagem de, se os *raw data* tiverem na sua origem o mesmo intervalo temporal e geográfico, a inclusão de ambos os *raw data* iria muito provavelmente permitir uma maior aproximação à realidade, em virtude de uma maior quantidade de informação sobre a observação de um artefacto no tempo e espaço. No caso da implementação que foi feita, embora seguindo esta última abordagem, como os *raw data* tem origens distintas tanto ao nível temporal como geográfico o *merge* de ambos os *raw data* não permite extrair mais-valias na análise de mobilidade. No entanto demonstra-se a aplicação geral quando temos *raw data* com dados muito semelhantes, mas obtidos de diferentes sensores onde é por isso necessário fazer a distinção ao nível do sensor. No caso da primeira implementação esta questão não se coloca porque cada sensor tem inclusivamente atributos próprios que os distingue, o GPS tinha a latitude e longitude, enquanto por exemplo, o Wi-Fi tinha o BSSID.

Recordando novamente a definição do conceito *Observation*, a representação do mesmo deverá ter em consideração que se observa um artefacto em determinada localização e instante temporal.

Assim e de forma similar à representação apresentada na Tabela 5.5, a estrutura de dados para a representação do conceito *Observation* é a seguinte:

Id_Observation	Artefact	Timestamp	Latitude	Longitude	Sensor
1	d96eca82bd	1383860930	45.49581987	9.1737967	Twitter
2	36280c79d4	1383860938	45.47882267	9.1866058	Twitter
100001	539270	1323314639	41.909303	-87.677131	foursquare
100002	1330946	1323320922	41.93530373	-87.647058	foursquare

Tabela 5.22 – Exemplo de *Observations* de redes sociais

No caso particular desta implementação o atributo *Location* é sempre geométrico e por isso subdividido em Latitude e Longitude. Foi ainda adicionado o atributo extra, *Sensor*, para a distinção entre observações de uma ou outra rede social.

De salientar que a transformação dos *raw data* da rede social *Twitter* foi muito simples com uma correspondência quase direta, contrariamente aos dados da rede social *Foursquare* onde foi necessário desenvolver um pequeno módulo de *software* para a associação dos dois *datasets* (*checkins*

e *venues*). A base da informação necessária para descrever uma observação estava na tabela de *checkins*, no entanto era necessário com base na identificação da *venue* extrair da tabela *venues* a informação geográfica das publicações realizadas. O resultado dessa associação deu origem então aos registos carregados para o conceito *Observation*.

Como os dados foram obtidos a partir de fontes de dados anteriormente utilizadas em trabalhos de investigação, não foi necessário muitos cuidados com os erros ou inconformidades relativamente aos registos e atributos associados. Neste caso em particular não foi necessário lidar com registos nulos ou falta de informação pelo que o processo de transformação ficou assim mais simplificado quando comparado com o mesmo processo de transformação da primeira implementação.

5.3.3. Lugares e Estadias

Após a conversão dos vários *raw data* em *Observation*, os processos de transformação seguintes visam a extração de dois conceitos: *Places* e *Stays*. Quando a monitorização dos artefactos é constante e realizada em pequenos intervalos de tempo, embora com maior ou menor dificuldade, é possível inferir se não todos, a maioria dos *places* mais relevantes para os indivíduos e/ou relevantes do ponto de vista global, e em simultâneo inferir também as *stays* e respetiva duração das mesmas. Mas, através das redes sociais, em particular através das publicações realizadas, não existe um seguimento contínuo do utilizador, uma vez que as publicações são feitas de forma mais ou menos esporádica, não sendo sequer obrigatório que se faça uma publicação. É muito natural que, por exemplo no *Twitter*, haja utilizadores que sejam mais seguidores de outros utilizadores (e como tal consumidores das publicações) do que propriamente produtores de conteúdos, sendo por isso natural a fraca frequência de publicações.

Estamos assim perante um constrangimento na extração de *Places* e *Stays*, e não será o único. O número reduzido de publicações diárias na rede social quando comparado com, por exemplo, um sensor Wi-Fi a recolher dados com uma frequência de um minuto, dificultaria sempre a correta extração destes dois conceitos. Mas mais preocupante é o tipo de conclusões que este tipo de observações permite extrair relativamente aos restantes conceitos. Pela definição dos conceitos *Place* e *Stay* muito dificilmente teremos utilizadores que publiquem com uma frequência tal que se possa considerar que esteve num local durante um determinado período temporal e como vimos anteriormente, a extração dos restantes conceitos depende muito dos resultados da extração de *Places* e *Stays*.

Vejamos de forma individual dois exemplos de cenários de utilização destas duas redes sociais. Começando pela rede social *Foursquare*, um cenário natural passa por um indivíduo chegar a um restaurante e alguns momentos após a sua chegada fazer um *checkin*.

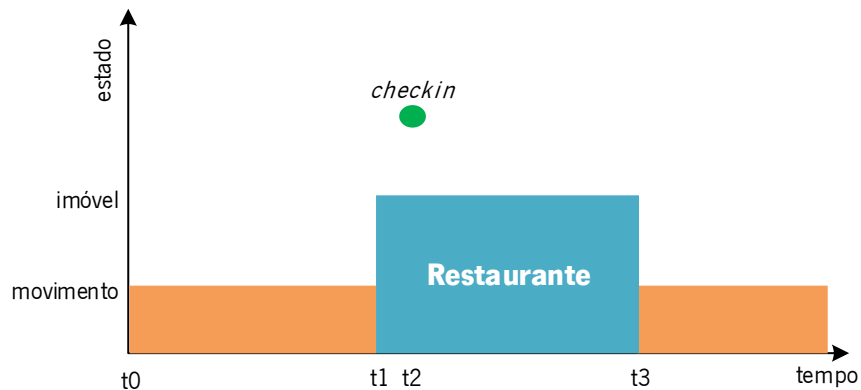


Figura 5.26 – *Checkin* num restaurante

Pela representação gráfica da Figura 5.26 que ilustra o cenário descrito anteriormente, o utilizador chegou ao restaurante no instante t_1 , fez o *checkin* no instante t_2 e abandonou o local no instante t_3 . Embora não sabendo se a diferença temporal entre t_1 e t_2 é mínima para que se possa considerar que o utilizador fez o *checkin* imediatamente após a chegada ao local, vamos assumir que sim e o t_2 poderia então ser o início da estadia nesse local. Não considerando então relevante a diferença temporal podemos afirmar que é possível saber o momento inicial da estadia, mas não temos forma de saber, apenas com os dados do *Foursquare*, o momento em que a estadia terminou, o instante t_3 . Faltando esta informação não é possível definir uma *Stay* e consequentemente determinar se um *Place* é ou não relevante para o indivíduo com base no tempo de presença nesse local. O melhor que poderemos inferir é que, embora não seja possível representar o conceito de *Stay*, poderemos assumir que seja possível considerar como *Place* relevante os locais mais visitados pelo utilizador. A questão que imediatamente emerge é o facto de no *Foursquare* os locais onde se realizam os *checkins* serem pontos de interesse, como por exemplo, restaurantes ou atrações turísticas; e nesse caso será muito difícil senão impossível identificar dois dos *places* mais relevantes de cada indivíduo: a sua habitação e o local de trabalho ou estudo. O local de trabalho ou estudo ainda poderá estar disponível nesta rede social, por exemplo, a Universidade do Minho está presente, mas a habitação de cada indivíduo não é um local típico de *checkin* no *Foursquare*.

Em conclusão, a partir das observações da rede social *Foursquare* o conceito *Stay* não será possível de extrair por falta de informação relativo ao fim da estadia e mesmo relativamente ao início da estadia existiria sempre um erro, mesmo que mínimo, correspondente ao instante do *checkin* e a

chegada ao local. Relativamente ao conceito *Place* poderemos considerar que os locais onde foi feito o *checkin* são locais relevantes para o indivíduo, mas com a ressalva que poderão estar em falta alguns dos mais relevantes.

No caso da rede social *Twitter*, vamos imaginar o seguinte cenário: um utilizador apenas publica *tweets* quando se encontra imóvel e como tal estará num *place* e enquanto esteve nesse *place* publicou em diferentes momentos dois *tweets*.

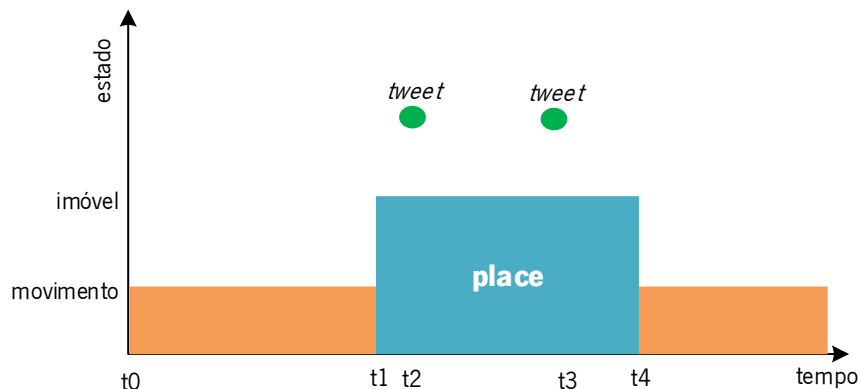


Figura 5.27 – Tweets no tempo

Contrariamente às publicações na rede social *Foursquare*, neste caso é muito menor a precisão em afirmar que o primeiro *tweet* realizado no instante t_2 corresponde ao início da estadia no local, sendo a diferença temporal entre t_1 e t_2 suficientemente pequena para que não seja possível afirmar que existiu movimento. Enquanto no *Foursquare*, embora existindo a hipótese de estarmos a cometer um erro de avaliação, poderemos dizer que a estadia se iniciou com o *checkin*, no caso do *Twitter* como o princípio da publicação de *tweets* é muito diferente do conceito de *checkin*, assumir *a priori* que o primeiro *tweet* que se realizou num local corresponde ao início da estadia nesse local poderá ter um erro grosseiro de avaliação. Além deste aspeto, e como mostra a Figura 5.27, a distância temporal entre o último *tweet* e o fim da estadia poderá também ser significativo. Este cenário não é mesmo assim dos piores, porque temos duas publicações e poderíamos afirmar, embora com pouca precisão, que a estadia se resumiu ao período entre as duas publicações, t_2 e t_3 respetivamente. Vamos agora imaginar que o segundo *tweet* não existiu ou que se realizou já noutro local: neste caso tudo se complica pois não existe qualquer base na informação existente nas observações que dê garantias quando à mobilidade do indivíduo. Torna-se por isso difícil, como acontece com a rede social *Foursquare*, afirmar com alguma garantia que existem *Stays* apenas com os dados recolhidos da rede social *Twitter*. Mais uma vez relativamente ao conceito *Place*, é arriscado assumir que os locais onde se realizaram os *tweets* correspondem aos mais relevantes para os indivíduos, uma vez que não existe

informação de *stays* que apoie a extração dos *places* com base no critério temporal (os locais mais relevantes serem aqueles onde o indivíduo passa a maior parte do seu tempo). O máximo que poderemos fazer, e conscientes que poderemos estar a cometer erros de avaliação, será novamente considerar como relevante os *places* com maior número de visitas, neste caso em particular, das redes sociais, os locais onde existem o maior número de publicações.

Como forma de explorar esta abordagem, utilizando o número de publicações como referência para a relevância de um *place*, foram realizadas visualizações para utilizadores individuais com publicações de diferentes dias. Um dos utilizadores analisados efetuou, para um período de cinco dias de análise, um total de 92 *tweets* os quais serviram de base para a produção de um *heatmap*. Este *heatmap* agregou localizações próximas até ao máximo de 50 metros e quanto maior fosse o número de publicações nessa localização a cor tenderia do vermelho para o azul (a geração do mapa de cores atribuiu à cor vermelho o menor valor, logo menor atividade, e à cor azul o valor mais elevado de atividade).

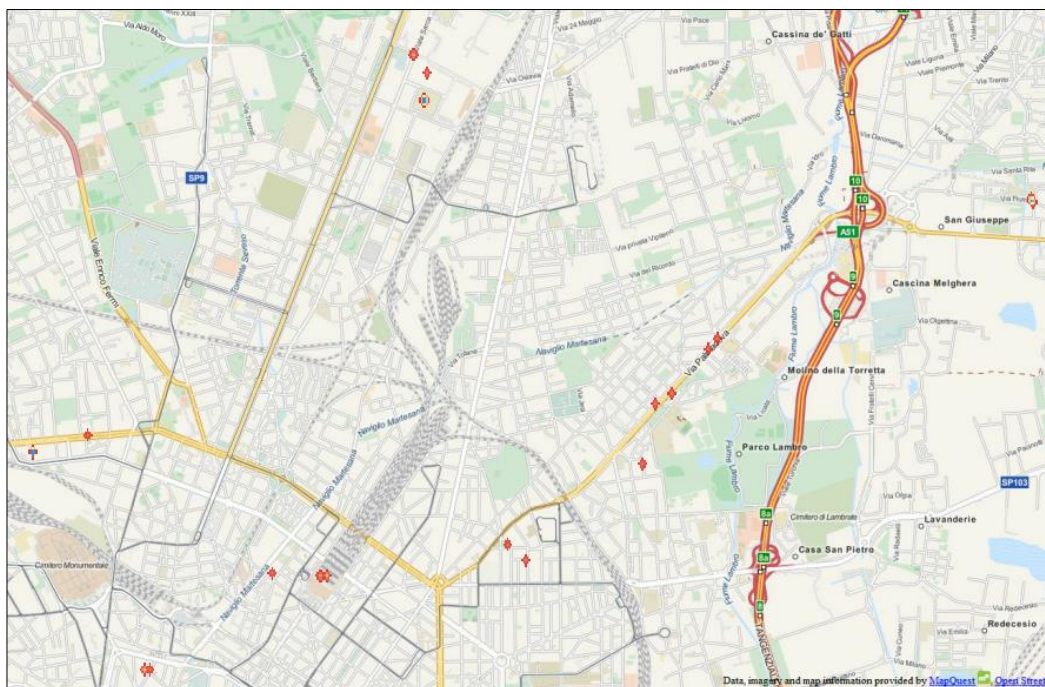


Figura 5.28 – Heatmap com as localizações dos *tweets* de um utilizador

A Figura 5.28 ilustra o resultado da produção da visualização no *qGIS*. Embora a imagem apresente uma área muito alargada de análise, dificultando a perceção dos locais onde foram publicados os *tweets*, é no entanto possível constatar que alguns deles coincidem com vias de transporte (linhas amarelas no mapa, maior detalhe na Figura 5.29b), sendo natural que os mesmos tenham sido feitos durante o movimento ou na fase inicial ou final dos movimentos. Logo aqui

verificamos então uma dificuldade nesta abordagem, pois temos um conjunto considerável de *places* que não correspondem na realidade a locais de relevância, mas sim a locais de passagem.



Figura 5.29 – Locais com maior ocorrência (a), locais coincidentes com vias (b)

A Figura 5.29a ilustra um dos locais com maior número de publicações (dos poucos locais assinalados a azul no *heatmap*). De qualquer forma apenas com esta informação disponível não é possível verificar se os *places* relevantes estão ou não representados e se efetivamente existem locais de passagem que desta forma são assinalados como *places*.

Com base nos resultados anteriores, a percepção sobre a utilidade dos dados das redes sociais parece cada vez mais ténue, sendo normal nesta fase colocar-se a questão se valerá ou não o esforço analisar os dados das redes sociais para a compreensão de padrões de mobilidade.

Se avaliarmos a mobilidade ao nível individual é extremamente arriscado extrair conclusões relativamente aos *places* relevantes, uma vez que existem dúvidas sobre quais os que são relevantes e o critério de como os extrair, no entanto, ao nível da análise global da mobilidade os dados são interessantes para a compreensão da ocupação do espaço. Vejamos a Figura 5.30 onde foram considerados, para um período de vinte e quatro horas, todos os *tweets* publicados por vários utilizadores (867 utilizadores, 2340 *tweets*). Na figura está apresentada apenas uma zona geográfica da cidade de Milão de forma a ser mais perceptível a distribuição da ocupação de espaço relativo a um dia de publicações.

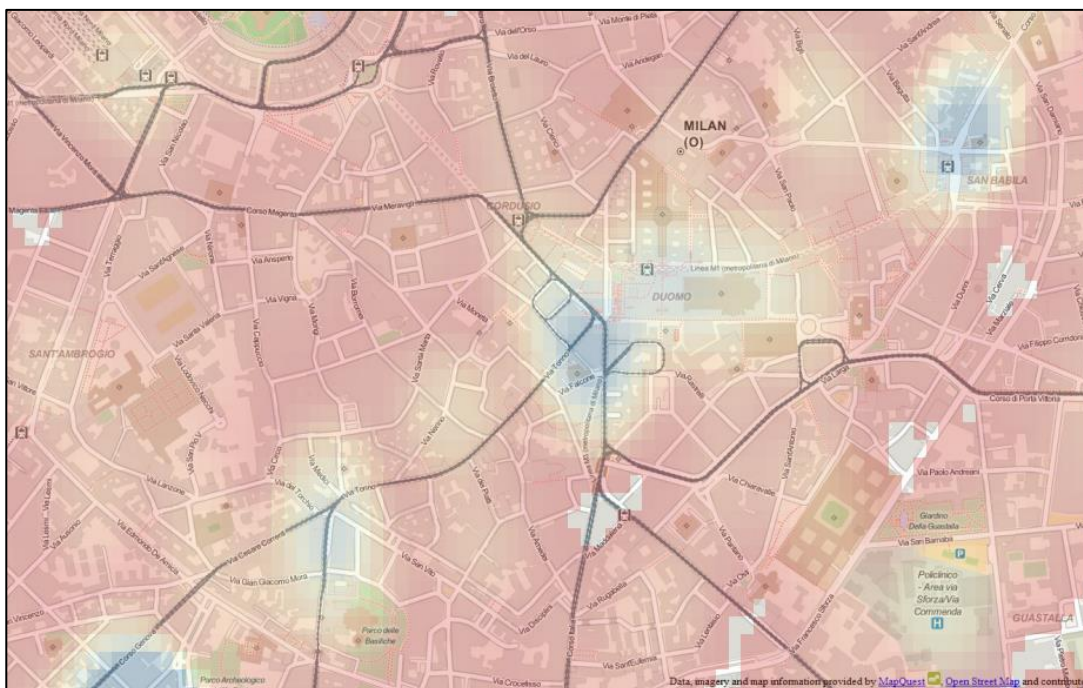


Figura 5.30 - Ocupação de espaço obtido através de 2340 tweets realizados por 867 utilizadores

Os locais com maiores ocorrências ao nível das publicações para um dia de análise, assinalados a azul na figura, obtiveram mais de 20 publicações para uma área de 50 metros. A distribuição da gama de cores do *heatmap* foi feita com base em 5 classes com a seguinte distribuição:

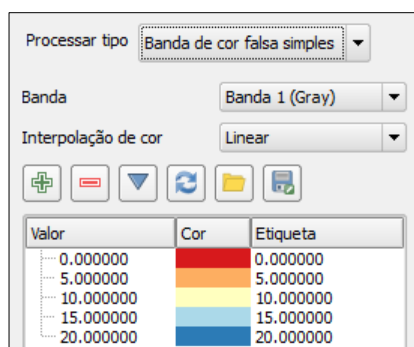


Figura 5.31 – Classificação de ocorrências por cores

Na Figura 5.32 foi realizada uma nova representação da ocupação de espaço para a mesma área geográfica, mas com menor sensibilidade ao nível do número de ocorrências. Neste caso os locais mais visitados são extraídos com um valor menor de ocorrências, 10 ou mais ocorrências, contrariamente ao exemplo anterior em que os locais mais visitados tinham que ter uma ocorrência igual ou superior a 20. Desta forma os locais mais visitados serão em maior número.

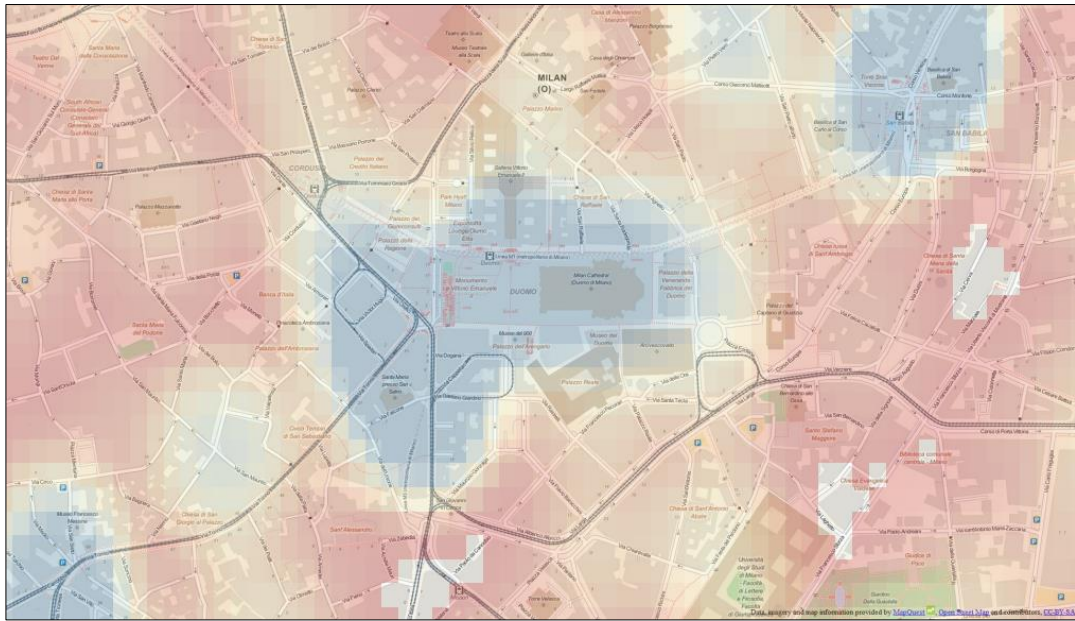


Figura 5.32 – Outra representação da ocupação de espaço com diferentes parâmetros de classificação

O resultado que se obteve através da produção de uma visualização com recurso a *heatmaps* demonstra a forma como os locais são mais ou menos visitados durante um dia. Diferentes visualizações poderiam ser feitas com a mesma base, apenas variando os períodos temporais de análise, seguindo exemplos existentes na literatura. Na Figura 5.33 temos um exemplo com maior detalhe de um local onde é expectável a existência de uma grande concentração de pessoas durante um dia de análise, um hospital. A área assinalada na figura com tons de azul reflete assim um número mais elevado de *tweets* comparativamente a outras zonas da cidade.

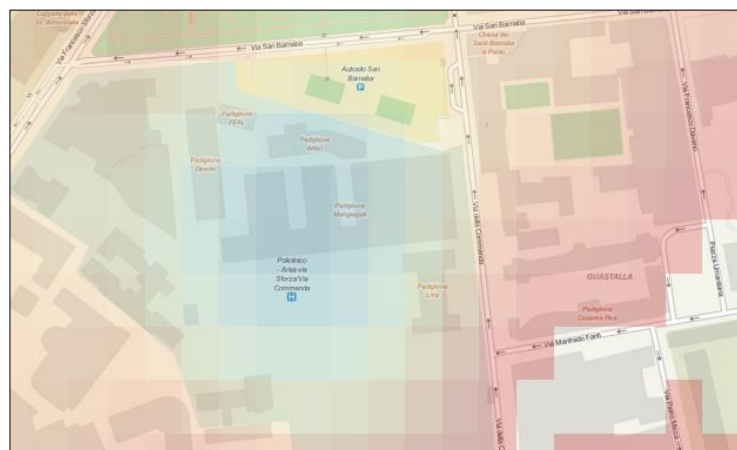


Figura 5.33 – Detalhe de uma zona geográfica

De acordo com a representação do conceito descrita na secção **Erro! A origem da referência não foi encontrada.** um *Place* é então descrito por um conjunto de observações e a

implementação desta representação para dados das redes sociais pode ser feita da seguinte forma, independentemente de ser a rede social *Twitter* ou *Foursquare*:

Id_Place	GPS		Lista de <i>Observations</i>
	Latitude	Longitude	

Tabela 5.23 – Estrutura de dados do conceito *Place*

Como um *Place* para esta implementação em particular apenas tem informação de GPS para caracterizar a sua localização, a estrutura de dados é muito semelhante à do conceito *Observation* pelo que o processo de transformação que se implementou foi relativamente simples. No caso da rede social *Twitter*, o processo de transformação poderia ser feito através de algoritmos de *clustering* espacial como anteriormente feito por vários investigadores. Neste caso, como tínhamos um modelo probabilístico fruto da implementação anterior, o que se fez foi recorrer ao mesmo modelo, mas de forma mais simplificada uma vez que apenas seria necessário a probabilidade P_1 . A opção por esta última abordagem prendeu-se apenas pela facilidade de utilização da implementação anterior e facilidade de adaptação para este novo cenário. No entanto para este cenário em particular, como o que se pretende é agregar apenas posições geométricas (não existindo posições simbólicas) que se encontrem geograficamente próximas, qualquer uma das abordagens seria sempre válida neste processo de transformação.

De salientar que para as visualizações apresentadas anteriormente com *raw data* do *Twitter*, a extração dos *Places* foi realizada umas vezes de forma individual (apenas os *places* por utilizador, filtrando as observações pelo atributo *Artefact*) e outras vezes de forma global, não sendo aplicado qualquer filtro ao atributo *Artefact*.

Com base na informação disponível no conceito *Place* e *Observation*, foram realizadas algumas visualizações de forma a serem testados os mesmos conceitos agora com dados obtidos da rede social *Foursquare*.

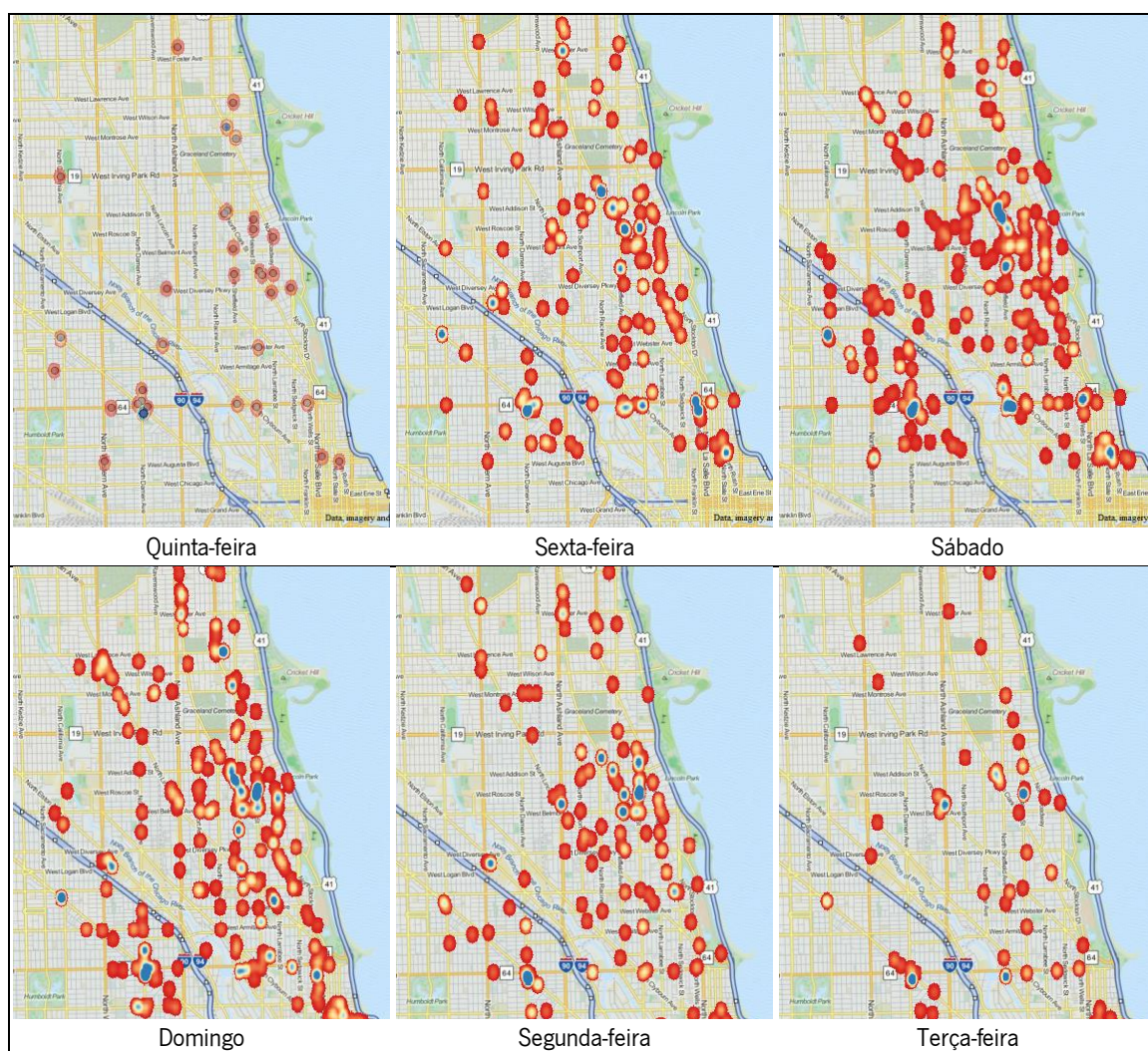


Figura 5.34 – Evolução da ocupação de espaço

A Figura 5.34, inspirada em trabalhos existentes na literatura, representa a ocupação do espaço urbano com base em *checkins*, para cada um dos dias da semana, num total de seis dias de análise, sempre para a mesma área geográfica e com todos os utilizadores disponíveis. O primeiro dia da figura corresponde a uma quinta-feira, o segundo a uma sexta-feira, e assim consecutivamente, ficando em falta um dia útil, a quarta-feira, que não é muito diferente do dia que a antecede e do dia que a sucede. A figura ilustra um fenómeno curioso: as maiores ocorrências verificam-se particularmente ao fim de semana; pelo que poderá ser resultado do contexto de publicação dos *checkins*. Tendo em consideração o contexto temporal e o facto de nesses dias existir maior atividade lúdica, é compreensível que aumente o número de *checkins* nesses dias, fruto das idas a espaços lúdicos e de confraternização, como por exemplo bares e restaurantes. A ser verdade esta hipótese aqui colocada, torna-se evidente que as conclusões extraídas apenas com recurso à rede social

Foursquare poderão ser enganadoras, uma vez que os *places* considerados como relevantes têm na sua maioria um carácter recreativo.

Apenas demonstrativa da hipótese apresentada, a Tabela 5.24 apresenta a relação entre o total de *checkins* que foram realizados num sábado e o período em que foram realizados: laboral e pós-laboral. Neste caso o que se verifica é que fora das horas normais de trabalho foram realizados mais de dois terços dos *checkins*.

Total de <i>checkins</i>	Horário laboral	Fora horário laboral
983	279	704
100%	28,4%	71,6%

Tabela 5.24 – Percentagens de *checkins* num sábado

Verificamos assim que mesmo através de vistas globais de mobilidade, neste caso particular focando a ocupação do espaço com recurso a dados obtidos das redes sociais, é necessário um extremo cuidado nas conclusões inferidas a partir das análises realizadas. No entanto, o mais importante para este trabalho não são propriamente as conclusões ou as formas de visualização relativas à mobilidade humana, mais sim a adequação dos conceitos de mobilidade e respetiva estrutura de conceitos que foi desenvolvida durante este trabalho.

Assim, quer com os dados do *Twitter*, quer com os dados do *Foursquare*, a estrutura de conceitos mantém-se ajustada e responde às necessidades de representação da informação. O que se verifica nesta segunda implementação relativamente à primeira é o fato de existirem conceitos que devido às características das redes sociais não chegam a ser utilizados, como é o caso do conceito de *Stay*. Por outro lado, na construção da estrutura de conceitos e validação com os dados obtidos através da monitorização via smartphone houve o cuidado de subdividir o conceito *Place* de acordo com a sua importância para o indivíduo ou para o grupo, daí termos a noção de *Place* candidato e *Place* relevante. Como com os dados das redes sociais é impossível a extração de *Stays*, a distinção entre estas duas tipologias de *Places* torna-se mais complexa, não pelo processo de transformação que lhe poderá dar origem, mas pela própria definição ou ideia subjacente à noção de relevante.

Identificaram-se pelo menos duas formas de considerar um *Place* como sendo relevante: ou pelo tempo de presença nesse local, que ao atingir um determinado valor temporal se considera o mesmo de relevante; ou pela frequência de visitas ao local e nesse caso são relevantes os locais onde os indivíduos vão mais bastas vezes quando comparados com os restantes locais visitados.

O problema que se coloca nas redes sociais é que no primeiro caso não se consegue estimar o valor temporal das estadias nos locais e no segundo caso não conhecemos todos os locais que foram visitados pelos indivíduos e como tal não se consegue estimar quais os locais com maior frequência de visita.

Por estes motivos, e de forma a evitarmos afirmações grosseiras ao considerarmos, por exemplo, um restaurante ou uma linha de Metro como locais relevantes para os indivíduos e não consideramos outros locais, com a sua habitação, como relevante, considera-se prudente que qualquer *Place* extraído de *Observations* de redes sociais, se usado de forma independente de outros sensores, sejam considerados apenas como *Places* candidatos.

5.3.4. Suspensão do movimento

Como referido na primeira implementação, o conceito *Time Leap* está presente quando não é possível extrair uma *Stay* devido à diferença temporal ser superior ao parâmetro T_{max} . Como os dados das redes sociais são esparsos temporalmente, não é possível a extracção de qualquer *Stay* pelo que existe assim uma forte probabilidade de ocorrerem *Time Leaps*.

Seria de esperar que no caso da rede social *Foursquare* os utilizadores apenas fizessem um *checkin* num determinado local e só voltariam a repetir esse mesmo *checkin*, se o repetirem, algum tempo depois, sendo natural que existam observações de *checkins* noutros locais nesse intervalo de tempo. No entanto, para surpresa, os dados mostram precisamente o contrário. Existem utilizadores que efetivamente têm observações consecutivas com o mesmo *place*.

A Tabela 5.25 ilustra os primeiros sete registos de um utilizador, de um total de vinte e sete registos, onde é possível verificar que o *place* identificado nas cinco primeiras e consecutivas observações é sempre o mesmo, estando por isso em presença de *time leaps*.

Timestamp	Latitude	Longitude	Place
09-12-2011 14:51:00	44.88354615	-93.21148396	1
10-12-2011 20:49:00	44.88354615	-93.21148396	1
11-12-2011 18:56:00	44.88354615	-93.21148396	1
12-12-2011 10:30:00	44.88354615	-93.21148396	1
12-12-2011 16:23:00	44.88354615	-93.21148396	1
14-12-2011 15:25:00	44.884415	-93.21308	2
14-12-2011 17:01:00	44.88354615	-93.21148396	1

Tabela 5.25 – Exemplos de observações no mesmo *place*

O processo de transformação implementado para a extração de *Time Leaps* é bastante simples e direto, em tudo semelhante ao processo de transformação desenvolvido na primeira implementação, processo este que a partir de duas observações consecutivas que referem o mesmo *place* extrai um *time leap*. A representação usada para esta implementação foi a mesma da implementação anterior, apresentada na Tabela 5.16. A forma e o conteúdo da representação deste conceito não varia de acordo com a tipologia do sensor, antes pelo contrário, é mantida inalterável. A Tabela 5.26 demonstra alguns exemplos dos *time leaps* extraídos para um indivíduo.

Id_TimeLeap	Place	Timestamp_Initial	Timestamp_Final	Leap_time
1	1	1323442260	1323550140	107880
2	1	1323550140	1323629760	79620
13	1	1327340940	1327517100	176160

Tabela 5.26 – Exemplos de *time leaps* extraídos

A existência de *time leaps* com valores tão elevados no atributo *Leap_time* é em certa medida sugestivo de que poderão estar em falta a identificação de outros locais que poderão ser de relevo para o utilizador. Como mostra a Tabela 5.25 os quatro primeiros registos demonstram *checkins* em quatro dias distintos, pelo que muito dificilmente este utilizador apenas esteve num único local nestes quatro dias. Por este motivo, ganha ainda mais força a recomendação de que as análises de mobilidade com base em dados de redes sociais devem ser feitas com muitas reservas.

5.3.5. Movimento

Seguindo o mesmo princípio do conceito *Stay*, para que o conceito *Elementary Movement* exista seria necessário que as publicações em redes sociais permitissem observar o indivíduo em intervalos de tempo muito curtos. No entanto, nem pelo objetivo, nem pelo funcionamento das redes sociais isso acontece. As redes sociais têm objetivos bem concretos e em nenhum deles se prevê a monitorização dos seus utilizadores, pelo menos é esse o comportamento esperado.

Assim, os dados recolhidos são fruto das publicações, e refletem os comportamentos não de mobilidade em si, mas de utilização da rede social por parte do indivíduo. Por este motivo nunca teremos dados obtidos em intervalos predefinidos de recolha, mas apenas quando existe uma publicação. Esta realidade implica que não seja possível ter duas observações consecutivas no tempo com uma distância tão curta que permita inferir a posição intermédia e como tal impossibilitando a extração do conceito *Elementary Movement*.

Embora não sendo possível extrair *Elementary Movements* a partir de observações de redes sociais, o facto de tanto os *raw data* do *Twitter* como do *Foursquare* possuírem informação associada à *Location*, permite a extração de *Space Leaps*. Mais uma vez, tal como acontecia na primeira implementação, os dados usados para a extração de *Space Leaps* referem-se ao período de movimentação.

A representação é exatamente igual à que foi apresentada na Tabela 5.18, com informação da *Location* inicial e final assim como do *Timestamp* inicial e final. A Tabela 5.27 ilustra três de um conjunto mais alargado de *space leaps* que foram extraídos para um utilizador em particular.

Id_SpaceLeap	Location_Start		Location_End		Timestamp_Initial	Timestamp_Final
	Latitude	Longitude	Latitude	Longitude		
5	44.883546	-93.211483	44.884415	-93.21308	1323706980	1323876300
6	44.884415	-93.21308	44.883546	-93.21148	1323876300	1323882060
8	44.883546	-93.211483	44.854609	-93.242069	1324667220	1324753620

Tabela 5.27 – Exemplo de três *space leaps* de um utilizador

Com base na extração dos *space leaps* a partir dos dados recolhidos da rede social *Foursquare* para um único utilizador, é facilmente possível, recorrendo por exemplo à ferramenta *Gephi*, fazer uma representação geográfica baseada em grafos. A Figura 5.35 ilustra exemplos desses *Space Leaps*, onde os arcos a verde representam movimentos e a cinza as *locations* de origem e destino dos movimentos. Como existem variações muito significativas nas coordenadas de GPS entre as *locations* um dos arcos associados a um *space leap* é significativamente mais extenso pelo que dificulta a leitura da figura. Por esse motivo foram adicionados à figura duas imagens que ilustram com mais detalhe as transições entre *locations* mais próximas (canto superior esquerdo e canto inferior direito da figura).

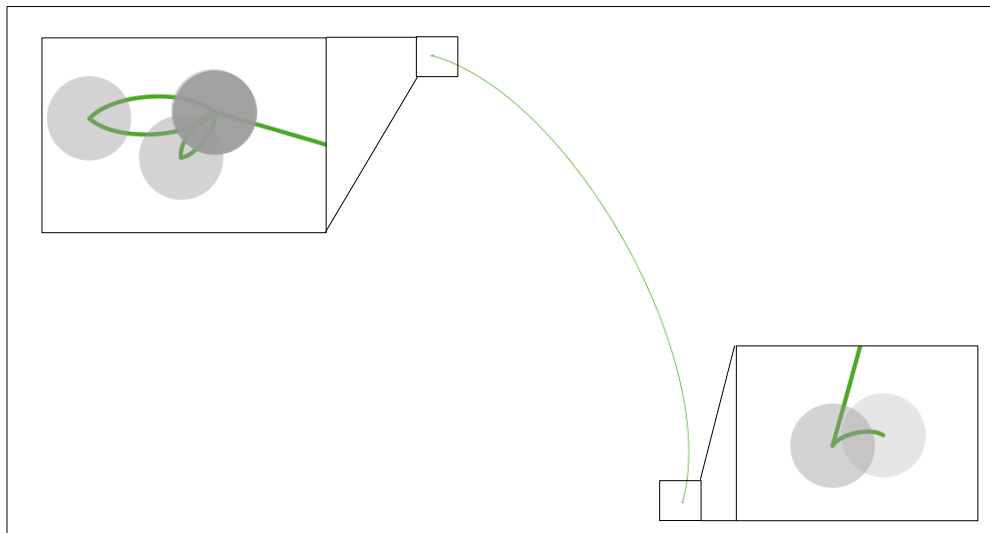


Figura 5.35 – Representação gráfica dos *space leaps*

Na Figura 5.35 podemos verificar um *space leap* entre duas *locations* mais distantes, no canto superior esquerdo destaque para quatro *space leaps* com distâncias relativamente reduzidas comparativamente com o *space leap* descrito anteriormente e no canto inferior direito um outro *space leap* novamente com uma distância reduzida.

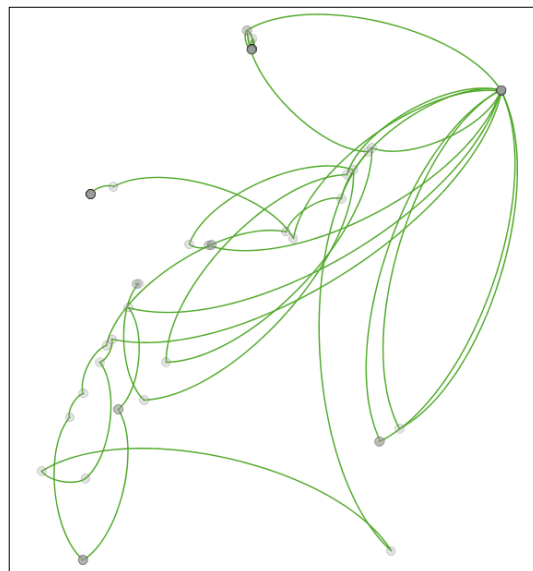


Figura 5.36 – *Space Leaps* de observações *Twitter*

De forma similar, foi também feita uma representação dos vários *space leaps* de um único utilizador da rede social *Twitter*, com informação recolhida entre 7 de Novembro de 2013 e 2 de Dezembro de 2013, aproximadamente um mês de observações. A Figura 5.36 apresenta a verde os vários *space leaps* e a cinza estão assinalados os *places*.

Se compararmos a Figura 5.36, que ilustra os *space leaps*, com a Figura 5.37, que ilustra os *places* desse utilizador, verifica-se uma sobreposição perfeita entre ambas as figuras. Com a combinação de ambas as visualizações é possível estabelecer relações entre os diferentes *places* visitados por este utilizador, o que analisadas de forma individual seriam mais difíceis de concluir. Na Figura 5.36 poderemos verificar com base nos arcos associados aos *space leaps*, que existem *places* que poderão ser mais relevantes com base na conectividade com os restantes *places*, por exemplo, no topo direito da figura existe um *place* com um número maior de conexões para outros *places* sugerindo que este possa ser relevante para o utilizador. Uma possibilidade para se validar se os *places* poderão ser relevantes para o utilizador seria observar os padrões temporais e espaciais associados a esses *places*.

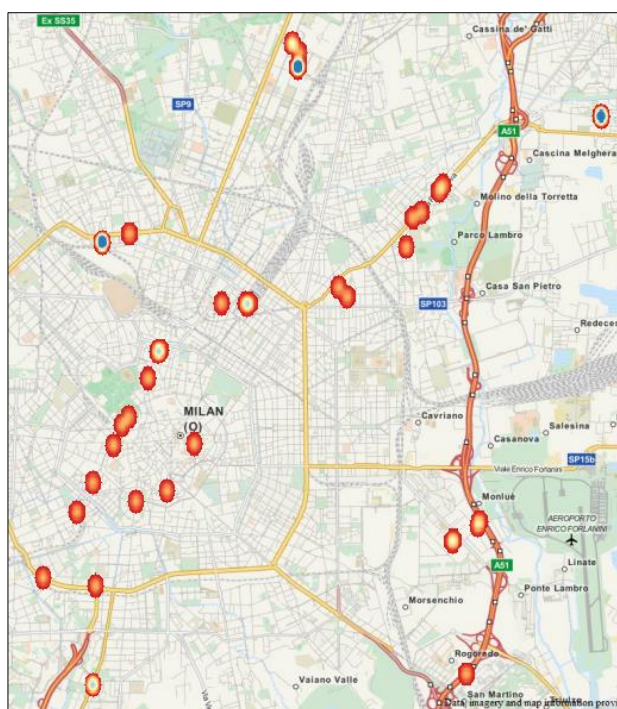


Figura 5.37 – Localização dos Places

Por exemplo, no topo da Figura 5.37 está um dos locais com maior número de publicações para este utilizador, e se considerarmos a contextualização espacial do local, indica que o utilizador realizou publicações no *Twitter* quando se encontrava no centro comercial Bicocca Village⁴² em Milão.

⁴² Informação sobre o Centro Comercial em: <http://www.bicoccavillage.it/> (visitado em Agosto de 2016)

A dúvida neste caso é saber se é um local muito visitado por ser cliente do centro comercial ou se é pelo facto do utilizador ser funcionário nesse centro, permanecendo a questão quanto à sua relevância para o utilizador. De qualquer forma com a contextualização espacial deste local já foi possível perceber o motivo de atração associado a este local.

Contrariamente ao que se verificou na primeira implementação, neste caso devido às fontes de dados usadas, não existem observações simbólicas, pelo que os *space leaps* serão todos eles geométricos.

5.3.6. Trajetória

Mais uma vez, para o final fica o processo de transformação associado ao conceito *Trajectory* e mais uma vez é em tudo idêntico ao que foi desenvolvido na primeira implementação. O conceito geral indica que uma *Trajectory* é uma sequência ordenada no tempo dos vários *Elementary Movements* e *Space Leaps* que ocorreram ao longo do intervalo de análise. Neste caso particular das redes sociais, como não existem *Elementary Movements*, o processo fica ainda mais simplificado, pois a *Trajectory* resume-se então apenas a *Space Leaps*.

Recorrendo novamente aos dados de um único utilizador do *Twitter* com publicações em diferentes dias consecutivos poderemos ver diferentes padrões de mobilidade com base na representação do conceito *Trajectory*, salvaguardando sempre que poderão estar em falta muitos dos movimentos realizados pelo utilizador, que por falta de dados de mobilidade não permite a sua representação nos diferentes mapas individuais de mobilidade.

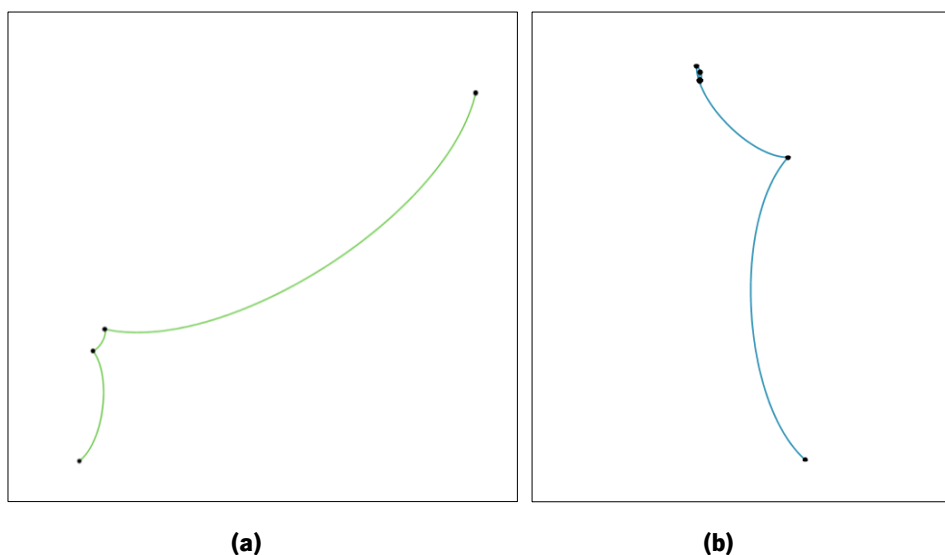


Figura 5.38 – Representação do conceito *Trajectory* em dois dias distintos para o mesmo utilizador

A Figura 5.38 ilustra dois dias consecutivos, terça (a) e quarta-feira (b) respetivamente, onde se verifica a estadia em diferentes locais da cidade de Milão e consequentemente diferentes movimentações no espaço urbano.

Este pequeno exemplo para um único utilizador demonstra ainda que se fizéssemos apenas uma análise com dados agregados e com esses produzíssemos mapas globais de mobilidade (semelhantes à Figura 5.36) não seria possível verificar que existem padrões comportamentais distintos inclusivamente para diferentes dias úteis da semana. É sempre expectável que existam padrões de mobilidade diferentes quando comparamos dias uteis com fins de semana, mas neste caso verificamos que mesmo entre diferentes dias uteis existem comportamentos distintos.

Esta constatação apenas foi possível de ser identificada porque foram gerados mapas individuais de mobilidade para cada um dos dias, o que permite reforçar a hipótese apresentada no Capítulo 3 - Figura 3.4, de que existem obvias mais-valias em gerar inicialmente mapas individuais de mobilidade e posteriormente fazer a sua agregação quando comparado com análises baseadas apenas em dados agregados e consequentes mapas globais.

5.4. Conclusões

Com o desenvolvimento da primeira implementação, descrita na seção 5.2, pretendeu-se por um lado validar e consolidar a estrutura de conceitos proposta e por outro lado criar um modelo computacional que demonstrasse a mais-valia da utilização de múltiplos sensores e a consequente fusão de dados de mobilidade. Pretendia-se, neste último caso, demonstrar que o recurso a diferentes dados de diferentes sensores, quando processados em paralelo, permitia uma melhor caracterização dos padrões de mobilidade individuais dos utilizadores. Assim, com base nos dados recolhidos pela aplicação *MOVE* e no modelo probabilístico criado para a extração dos dois principais conceitos da estrutura de conceitos, *Place* e *Stay*, realizou-se uma experimentação que consistia na adição incremental dos dados dos vários sensores disponíveis.

Assim, o primeiro teste teve apenas em conta as observações obtidas a partir de dados GPS (e consequentemente relacionados com a probabilidade P_1 do modelo probabilístico). Apenas com um sensor verificou-se que para um dado utilizador em particular, dos dez locais identificados no diário para uma semana de análise (o mesmo utilizador usado para documentar os resultados na implementação), apenas foi possível identificar como *places* relevantes, dois desses locais.

Posteriormente foram consideradas também as observações Wi-Fi, mas tendo em conta nesta fase apenas a semelhança entre os endereços MAC (correspondente à probabilidade P_4). O processamento simultâneo dos dados destes dois sensores permitiu uma evolução na identificação dos *places* relevantes, passando agora a três.

No teste seguinte foram incluídas as probabilidades que associam a proximidade temporal entre observações GPS e Wi-Fi (P_2). Com a inclusão da componente temporal na análise, os resultados obtidos melhoraram e passou a ser possível identificar cinco *places* relevantes, dos dez descritos no diário do utilizador.

Finalmente no último teste realizado, foram incluídas todas observações dos três sensores disponíveis e respetivas probabilidades associadas, procedimento usado na experimentação descrita na seção 5.2. Neste último teste verificou-se ser possível identificar seis dos dez *places* possíveis. A Tabela 5.28 resume os resultados obtidos da experimentação realizada.

Probabilidade	Diário	<i>Places</i> extraídos
P_1	10	2
P_1 e P_4	10	3
P_1 , P_4 e P_2	10	5
Todas	10	6

Tabela 5.28 – Resultados da fusão de dados

A Tabela 5.28 ilustra o incremento ao nível da deteção de *places* (que originou também um incremento ao nível das *stays*) com a adição incremental de mais fontes de dados no algoritmo de *place learning*. O objetivo desta experimentação visou a demonstração de que com a fusão de diferentes sensores no algoritmo de *place learning* obteríamos melhores resultados comparativamente com a análise baseada apenas num único sensor.

A utilização, neste caso em particular, de três tipos diferentes de sensores (podendo facilmente ser possível aumentar este número, bastando para isso a adição de mais probabilidades ao modelo e respetivos processos de transformação) e o seu processamento em simultâneo permitem uma extração de *Places* e *Stays* mais aproximada da realidade apresentada através dos registos presentes nos diários. Como todas as restantes perspetivas de movimento dependem desta extração, a fusão de diferentes sensores, ao melhorar a qualidade dos resultados ao nível de *Places* e *Stays*, influencia assim os resultados dos restantes conceitos e a consequente construção de um mapa de mobilidade, quer esta seja individual ou global, mais coincidente com a realidade.

Nas duas subsecções anteriores foram realizadas experimentações com diferentes tipologias de dados de forma a estudar/validar o potencial da estrutura de conceitos proposta no **Erro! A origem da referência não foi encontrada.** Contrariamente à primeira experimentação que foi realizada ao longo do desenvolvimento e evolução da referida estrutura, a segunda experimentação foi realizada já após a finalização desse desenvolvimento, tendo em mente a avaliação do potencial de utilização da estrutura de conceitos em diferentes tipologias de dados sem a necessidade de existirem significativas modificações à estrutura e conceitos. Por ser uma experimentação pós-definição da estrutura existiam naturais questões sobre a adequabilidade da mesma a novas tipologias de dados, neste caso em particular, obtidos a partir de duas diferentes redes sociais. No entanto os resultados obtidos demonstram que a implementação dos processos de transformação dos dados e a adequação dos conceitos foi de modo geral relativamente natural, não exigindo nenhuma modificação na estruturação dos conceitos nem a adição de nenhum novo conceito, pelo que a estrutura proposta demonstra-se assim ajustada tanto para a análise de mobilidade a partir de dados recolhidos de forma individual por múltiplos sensores, como também a outras tipologias de dados comumente usadas na análise de mobilidade humana.

Antes de se iniciar a segunda experimentação existia uma forte suspeita de que tendo em conta as características dos dados das redes sociais e o comportamento associado à publicação nestas mesmas redes sociais, era perfeitamente natural que alguns dos conceitos não fossem contemplados nesta experimentação. Os resultados demonstraram que, de forma geral, com base em dados oriundos de redes sociais, conceitos como *Stay* e *Elementary Movement* não têm correspondência com a informação disponível e como tal não poderão ser implementados. A Figura 5.39 apresenta a estrutura de conceitos de acordo com as necessárias adequações por força dos *raw data* das redes sociais, onde a cinza se ilustra os conceitos e processos de transformação que não foram implementados devido às características dos dados.

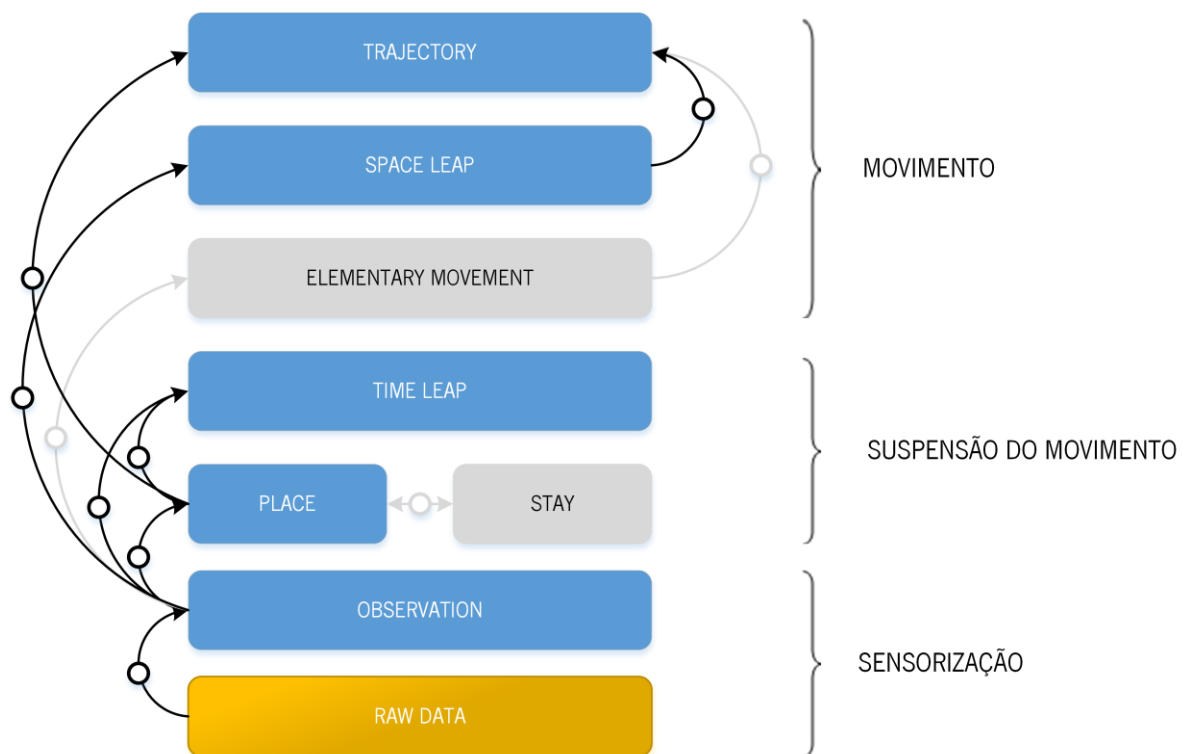


Figura 5.39 – Estrutura de conceitos ajustada às redes sociais

Os resultados da experimentação demonstraram ainda que os dados utilizados na primeira experimentação permitem a distinção entre os dois tipos de *Places* (relevante e candidato) com a definição de critérios que elevam os *places* candidatos a *places* relevantes, em particular tendo em consideração o tempo associado às *Stays*. No caso da segunda experimentação, em virtude de ser dificultada a extração do conceito *Stay*, o critério utilizado na primeira experimentação não se pode aplicar. Como acontece com a implementação dos vários processos de transformação, também os critérios usados nas várias implementações devem ser ajustados e modificados de acordo com o cenário aplicacional em questão. Embora com menor segurança do que aconteceu com a primeira experimentação, para o caso das redes sociais poderemos considerar o número de visitas a um *place* como um potencial critério de distinção entre relevante e candidato. Seriam *places* relevantes os locais mais visitados durante um período temporal. A insegurança desta afirmação e respetiva validade do critério acontece porque poderão ocorrer situações em que as publicações nas redes sociais não são feitas nos locais mais relevantes para os indivíduos (casa, trabalho, entre outros), por exemplo, um indivíduo poderá estar a publicar *tweets* enquanto viaja de transportes públicos e nesse caso as paragens associadas aos transportes públicos poderiam ser considerados de relevantes ou ainda fazer uma publicação no *Foursquare* nos locais que visita socialmente, como um restaurante, mas para o

indivíduo o que poderá ser relevante seria a sua casa ou o seu local de trabalho que são ainda mais visitados que esse restaurante mas que não são alvo de publicação na rede social *Foursquare* e como tal não são considerados.

Estas considerações quanto às conclusões que se extraem da análise de mobilidade com recurso a redes sociais vão ao encontro da salvaguarda feita anteriormente na seção 3.1. É aconselhável que as conclusões extraídas das análises de mobilidade com base em redes sociais sejam feitas com algum cuidado uma vez que as mesmas poderão estar desfasadas da realidade. Com base nos dados extraídos das redes sociais é possível encontrar alguns dos locais visitados pelos indivíduos, no entanto, devido às características da rede social e do comportamento de publicação na mesma, assim como o carácter esporádico com que se realizam as publicações, não é possível afirmar com um grau de certeza elevado que determinado indivíduo visita e permanece num local num período temporal considerável para que esse local possa ser relevante para este e que os movimentos que realiza possam ser extrapolados a partir dos locais identificados. As representações feitas na Figura 5.35 e Figura 5.36 ilustram apenas os *Space Leaps* que foram possíveis de identificar com os dados disponíveis, mas que muito possivelmente não identificam todas as mudanças de locais que os indivíduos realizaram durante o período de análise. Levando ao extremo, para ser possível através das redes sociais identificar a totalidade de mudanças de locais (e consequentemente, a existência de movimento) era necessário que os utilizadores realizassem pelo menos uma publicação em todos os locais que visitaram ao longo do dia. Como não é facilmente possível assegurar que este pressuposto é cumprido, as análises de mobilidade humana com base em redes sociais deverão ser consideradas meramente indicadoras de potenciais fenómenos de mobilidade existentes no espaço urbano. Assim, as mesmas deverão ser usadas com extremo cuidado e sempre que possível complementadas por outras análises de forma a atestar a sua coerência.

De salientar que o alerta para o uso cuidadoso dos dados de mobilidade recolhidos das redes sociais não implica que no futuro os mesmos não sejam considerados como uma hipótese de tipologia de dados na realização de estudos de mobilidade humana. Estes dados, usados amplamente em diversos trabalhos de mobilidade humana, foram e deverão continuar a ser opção, até pela facilidade com que os mesmos podem ser recolhidos em grandes quantidades, desde que as análises sejam feitas com algumas reservas. Inclusivamente, dependendo da forma como a recolha é realizada, é possível ter dados individuais e/ou dados agregados. No caso de existirem dados individuais, como aconteceu com os dados usados na segunda experimentação, é possível produzir mapas individuais

de mobilidade e posteriormente agregar estes mapas e gerar mapas globais, tirando partido das vantagens destas duas representações.

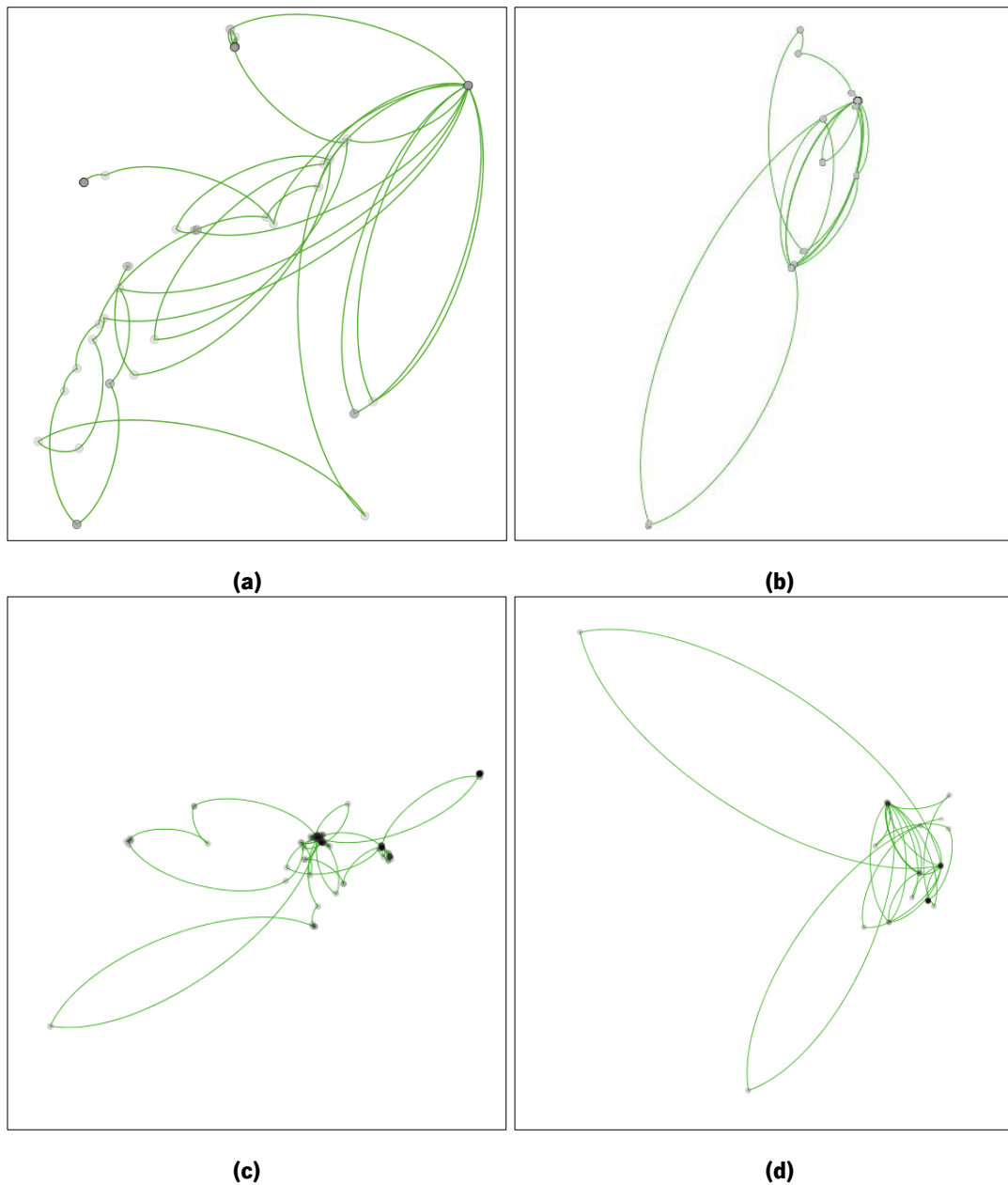


Figura 5.40 – Quatro perfis de mobilidade de quatro utilizadores para o mesmo referencial espaçotemporal

A Figura 5.40 apresenta quatro perfis de mobilidade com base nos dados recolhidos pela rede social *Twitter* para o mesmo referencial espaçotemporal, isto é, os dados foram recolhidos durante o mesmo intervalo de tempo e restritos à cidade de Milão.

Com estes quatro exemplos podemos verificar que: existem movimentações longas com menor frequência que as movimentações mais próximas (d); que existem padrões onde as visitas estão restritas a um número menor de locais mas com maiores ocorrências (c); que existem utilizadores onde as visitas são feitas a locais mais dispersos (a), enquanto outros utilizadores realizam visitas a locais mais próximos entre eles (d); entre outras conclusões que poderíamos extrair a partir dos vários mapas de mobilidade individuais.

Se realizarmos agora a agregação dos quatro mapas individuais num único mapa global de mobilidade o resultado é o seguinte:

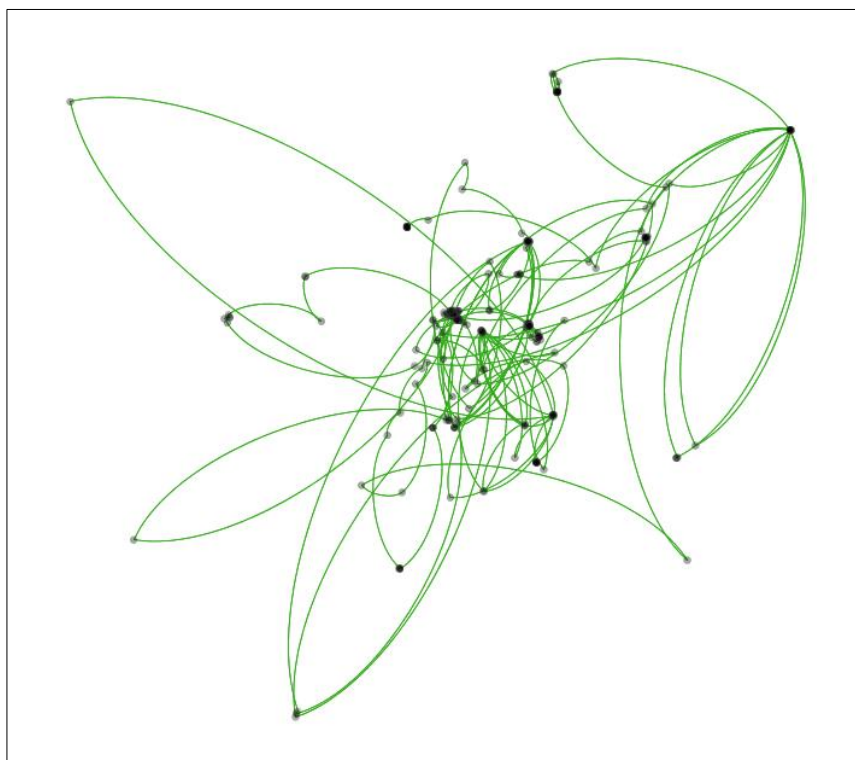


Figura 5.41 – Mapa global após agregação dos mapas individuais de mobilidade

Com a agregação da informação dos diferentes mapas individuais num único mapa global, obtemos uma visão mais rica para descrever a mobilidade dos artefactos ao longo do tempo e do espaço. No entanto, e como reflexo da tese apresentada na Figura 3.4, tanto os mapas individuais como os mapas globais têm vantagens e desvantagens, sendo por isso defendido como importante a existência de ambos. Se por um lado, a apresentação da informação apenas sobre a forma agregada tem o inconveniente de poder esconder determinados detalhes nos perfis individuais de mobilidade que são apenas possíveis de identificar através dos mapas individuais, por outro lado, a agregação dos

mapas de mobilidade permitem de forma simples a identificação de zonas territoriais que devido à densidade de população poderão ser pontos críticos de mobilidade.

A concentração de quantidades significativas de indivíduos numa área geográfica poderá ter reflexos por exemplo: no planeamento de vias de comunicação e serviços de transporte público, entre outras implicações. Se durante um período de análise como o que está representado na Figura 5.42 os locais assinalados a azul foram alvo de maior número de publicações e como tal (mais uma vez com algumas reservas) serem considerados por isso de locais relevantes para a análise que estamos a realizar, deverá então ser dada maior atenção a estes locais pois a movimentação destes indivíduos de e para esses locais poderá estar na origem de fenómenos nefastos para o espaço urbano.

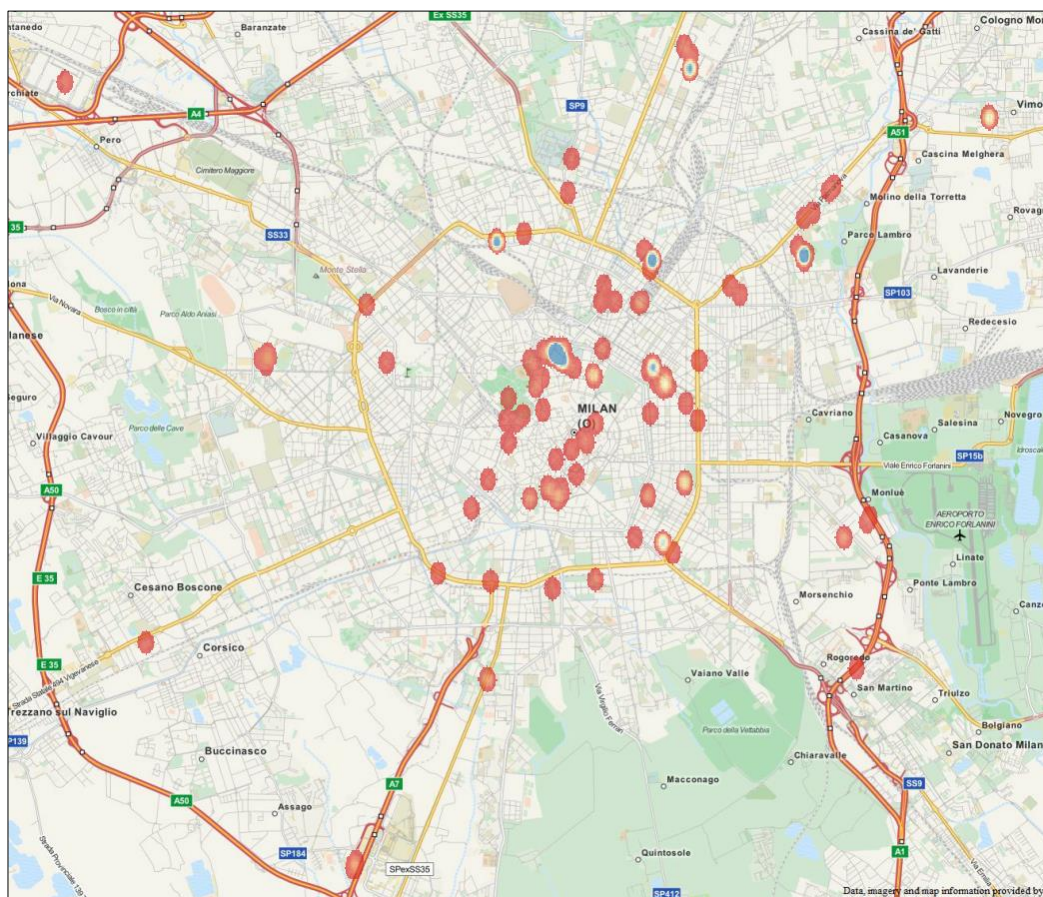


Figura 5.42 – Locais com maior número de publicações

Relativamente ainda à questão de os locais serem relevantes e para quem são efetivamente relevantes, a combinação de mapas individuais e mapas globais de mobilidade permitem identificar e caracterizar dois tipos distintos de locais relevantes. Enquanto com os mapas individuais de mobilidade, semelhantes aos apresentados na Figura 5.40 conseguimos identificar um conjunto de locais que pela frequência de visita podem ser considerados de relevantes para o indivíduo, já com os mapas de

mobilidade globais da Figura 5.41 ou com a representação dos locais mais visitados da Figura 5.42 conseguimos identificar os locais que poderão ser considerados de relevantes para o nosso espaço urbano.

A extração destes dois tipos distintos de locais relevantes e consequente destrição entre o que pode ser um local relevante para um indivíduo e um local relevante para o espaço urbano apenas se consegue obter seguindo uma sequência de etapas semelhantes às que se encontram descritas na Figura 3.4, produção de mapas individuais inicialmente e posterior agregação destes num mapa global com ou sem adição de mais mapas globais provenientes de outros sensores ou análises.

Capítulo 6. Conclusões

6.1. Ecosistema Urbano

O trabalho apresentado e descrito ao longo deste documento foca aspetos relativos à mobilidade humana e não podemos nem devemos pensar a mobilidade de forma isolada, uma vez que a mesma depende e influencia o meio envolvente. Os congestionamentos de tráfego, por exemplo, existem porque um conjunto alargado de indivíduos se movem num mesmo espaço durante um mesmo período temporal, mas também não deixa de ser verdade que se existem congestionamentos é porque a necessidade de mobilidade é muito superior à oferta existente. Portanto as duas faces do problema não vivem desassociadas, antes pelo contrário.

Focando em particular neste problema, o congestionamento de tráfego, a solução tem passado recorrentemente não pela resolução do problema, mas sim pelo remedeio do problema. Ano após ano, o procedimento é aumentar as vias de comunicação na tentativa de minimizar o problema, mas como não existe uma verdadeira identificação da causa do problema, ou o efeito não resultou como desejado ou originou um problema agora num outro ponto do espaço urbano. Este é um dos motivos que explica a constante necessidade de alterações às vias de comunicação e a continuidade de problemas em algumas das vias de comunicação.

Por este motivo, e porque nada de relevante se extrai de uma análise de mobilidade se não existir intervenção, foi incluído neste trabalho um capítulo (capítulo 3) direccionado apenas para esta temática, a definição do que poderá ser o Ecosistema Urbano. A inclusão deste tema revela-se importante no sentido em que todo o esforço que os investigadores têm despendido na produção e compreensão de padrões de mobilidade só faz sentido se aplicados os resultados no espaço urbano. Não podemos, nem devemos continuar a estudar padrões de mobilidade com os custos associados a

essa análise, para que depois quem planeia ou decide sobre o espaço urbano não os usar ou nem sequer ter conhecimento que os mesmos existem. Torna-se por isso importante envolver estes atores em todo o processo de análise de mobilidade humana e torná-los cada vez mais uma parte ativa e fundamental no processo para que os vários estudos que venham a ser realizados estejam alinhados com as necessidades do espaço urbano e dos decisores, e que as decisões sejam tomadas cada vez mais de forma consciente e baseada em factos ou conclusões o mais recentes possível.

6.2. Estrutura de Conceitos

A distância que neste momento ainda existe entre quem estuda os padrões de mobilidade e quem decide sobre o espaço urbano é muito considerável. No entanto, a consciencialização para essa realidade e a criação de iniciativas onde o cidadão é parte ativa na sua cidade irá com certeza aproximar estes dois atores.

Mas o aproximar e o incremento da interação entre equipas multidisciplinares que planeiam, decidem, intervêm e avaliam o espaço urbano, embora seja um grande contributo, não acrescenta muito à resolução dos problemas urbanos se não existir informação que permita a extração de conclusões sobre dinâmicas humanas.

Por esse motivo é necessário recolher e analisar elevadas quantidades de dados com correspondência direta à mobilidade humana, que poderão ser de apenas um sensor ou de um conjunto mais alargado de sensores.

À medida que se adicionam sensores, de forma a enriquecer a análise, adicionamos também maior complexidade quer ao nível da recolha, quer ao nível do processamento dos dados e extração de informação. Conscientes deste facto, a estrutura de conceitos desenvolvida e apresentada neste trabalho visa em primeira instância contribuir para a resolução destas dificuldades. Homogeneizar a informação de mobilidade, com diferentes conceitos para diferentes fenómenos associados à mobilidade humana, tendo por base a multiplicidade de sensores e respetiva multiplicidade de representações de dados.

Os fenómenos relativos a movimentos e imobilizações dos artefactos ao longo do tempo necessitavam de uma estrutura de conceitos que permitissem, por um lado, o suporte adequado à informação que se deseja representar e, por outro lado, garantir que os conceitos possuem a informação básica facilitadora da visualização gráfica desses mesmos fenómenos.

Assim a estrutura de conceitos desenvolvida cumpre com os requisitos exigidos para a representação de fenómenos espaçotemporais associados à mobilidade humana, relativos ao movimento e suspensão do movimento, e como demonstrado pelas duas implementações que foram realizadas, suficientemente completa e flexível para a representação de tipologias de dados muito distintas, quer ao nível da frequência e forma de recolha, quer ao nível dos comportamentos associados aos artefactos que estão a ser observados.

Outro resultado importante obtido pela definição da estrutura de conceitos prende-se com a representação proposta para o conceito *Observation* que permite unificar de forma genérica a representação das localizações simbólicas e geométricas dos artefactos.

Como nas duas implementações que foram realizadas era recomendável a realização de visualizações das diferentes representações de dados dos diferentes conceitos, uma vez que uma imagem demonstra melhor o resultado do que uma descrição textual ou tabular, a criação destas visualizações permitiu demonstrar também que a estrutura de conceitos apresenta-se adequada e de fácil integração em diferentes ferramentas de visualização. As três ferramentas utilizadas: *Gephi*, *qGis* e *Processing* têm pressupostos e objetivos muito diferentes e a sua utilização permitiu concluir que a estrutura idealizada para a estrutura de conceitos permite a produção simples de visualizações de acordo com a informação presente nos diferentes conceitos. Foi possível com base na informação disponibilizada através da estrutura de conceitos visualizar tanto a ocupação de espaço como conceitos associados ao movimento.

6.3. Fusão de dados

A construção da estrutura de conceitos e suas implementações realizadas, quer dos próprios conceitos, quer dos respetivos processos de transformação, além de terem como principal objetivo a validação e evolução dos conceitos, tinha como objetivo secundário a demonstração das vantagens da utilização de múltiplos sensores para a monitorização da mobilidade dos artefactos, quando comparada com a utilização de apenas um sensor.

Ao adicionarmos sensores ao processo de recolha estamos a adicionar também um custo extra, quer ao nível da própria recolha, quer ao nível do espaço de armazenamento. Por este motivo, era importante avaliar se a adição de um ou mais sensores justificaria o ganho nos resultados relativamente ao custo da sua inclusão. Se os resultados que se obtêm com um único sensor forem

muito idênticos aos resultados de múltiplos sensores, o custo extra no processo de recolha e processamento da informação não se justifica. A Tabela 5.28 apresentada na secção 5.4 demonstra que existe significativa e importante vantagem em se realizar a fusão de dados oriundos de múltiplos sensores. A fusão de dados torna-se ainda mais importante quando temos situações em que os mecanismos de recolha de dados têm algumas falhas e por isso existem lacunas nos dados recolhidos. Se por ventura não for possível recolher dados de GPS, por exemplo pela presença dentro de um edifício, esta lacuna poderá ser colmatada com as observações dos restantes sensores.

A extensão do algoritmo de *place learning* para acomodar outras fontes de dados, diferentes das que foram usadas nas duas implementações, apenas requer que sejam definidas as funções de probabilidade que avaliem a proximidade temporal e espacial entre um par de observações recolhidas no mesmo local.

6.4. Dados em bruto

Relativamente ao *raw data* utilizados para as duas implementações que foram realizadas verificou-se que conseguimos ter possivelmente os dois extremos: a monitorização constante dos artefactos através da aplicação *MOVE* instalada nos *smartphones* e, por outro lado, dados obtidos das publicações nas redes sociais realizadas de forma esparsa no tempo. Se no caso da primeira situação com a quantidade de observações que são recolhidas é possível rastrear quer as estadias quer as movimentações dos artefactos, no caso da segunda situação apenas se consegue afirmar que em determinado momento o artefacto foi visto num determinado local. No entanto independentemente do *raw data* verificamos que a estrutura de conceitos é alheia à forma ou mecanismo de recolha, lidando bem com ambas as situações.

Quanto aos resultados que se obtêm a partir de *raw data* das redes sociais, o que se conclui é que teremos que ter algum cuidado nas análises e interpretação dos resultados. Não nos podemos esquecer que os dados das redes sociais refletem preferencialmente comportamentos sociais e não propriamente comportamentos de mobilidade. Poderá ser sempre um bom ponto de partida para a análise de ocupação do espaço e até mesmo da mobilidade apenas com recurso ao conceito *Space Leap*, mas mesmo assim os resultados nunca serão vinculativos. Verificou-se com alguns dos resultados obtidos, por exemplo da rede social *Foursquare*, que os *check-ins* eram feitos preferencialmente em locais lúdicos e em períodos noturnos e fins de semana. Este comportamento

social nunca será um bom indicador de mobilidade humana numa cidade, quanto muito meramente indicativo de zonas de diversão noturna, por exemplo.

Por outro lado, a utilização de *raw data* idênticos aos que foram recolhidos pela aplicação *MOVE* também têm em si os seus inconvenientes. Neste caso em particular, o primeiro deles é o custo energético para os dispositivos móveis. Quanto mais sensores estiverem ativos e em pleno funcionamento nos *smartphones*, maior será o consumo de energia e como tal menor a autonomia dos equipamentos, originando inconvenientes para os proprietários dos equipamentos. O segundo inconveniente é que mesmo uma monitorização constante e cuidadosa irá ter sempre anomalias na recolha de dados, existindo períodos em que não existem observações recolhidas. Basta acontecer uma falta de energia e teremos situações idênticas às que foram identificadas na fase de validação da estrutura de conceitos. No entanto e optando por outras alternativas de recolha de dados onde as duas questões anteriores sejam mínimas, a monitorização constante de um artefacto poderá originar um inconveniente de igual ou maior amplitude comparativamente aos anteriores, as dimensões dos *datasets*.

Realizar o processamento de grandes *datasets* origina imediatamente duas questões tecnológicas: o espaço de armazenamento necessário para guardar tanto os *raw data* como depois as observações, e as necessidades de poder computacional para a realização dos processos de transformação sobre os dados recolhidos. Como estamos a trabalhar com dados individuais sem nenhuma agregação (o que seria também um inconveniente para os objetivos que pretendemos atingir), os processos de transformação necessitam de manipular e calcular um conjunto muito alargado de registos. A título de exemplo, cada utilizador da aplicação *MOVE* gerou em média cerca de 150 mil registos mensalmente, se pretendêssemos produzir uma vista global de mobilidade de um mês para um conjunto de 50 utilizadores, muito facilmente atingiríamos os 7 milhões de registos. Na implementação que foi realizada, em particular no algoritmo de *place learning*, a opção pelo modelo probabilístico como alternativa às técnicas tradicionais de *clustering* visou precisamente a otimização e criação de um algoritmo eficiente de processamento de observações.

6.5. Resultados das implementações dos conceitos

Começando pela segunda implementação, direccionada para a validação dos conceitos e respetiva estrutura com base em dados oriundos das redes sociais, verifica-se que os conceitos se

adequam na perfeição aos dados espaçotemporais recolhidos, onde a localização do artefacto é baseada exclusivamente em observações GPS. Verificou-se ainda que existe conceitos, que devido às características dos dados recolhidos, não são sequer utilizados. Tal facto não demonstra uma fraqueza da estrutura de conceitos, antes pelo contrário, só demonstra que a mesma está apta a coabitar com diferentes tipologias de dados, principio básico imposto neste trabalho aquando do desenvolvimento do mesmo.

No caso da primeira implementação, com dados obtidos de múltiplos sensores e onde a definição da localização do artefacto tanto poderia ser simbólica como geométrica, verificou-se que os conceitos permitem representar os diferentes fenómenos associados à mobilidade humana, quer ao nível da movimentação dos artefactos, quer ao nível da suspensão do movimento. Ficou a faltar nesta validação a existência de mais *datasets* de diferentes utilizadores e respetivos diários para melhorar a resposta dos parâmetros associados ao algoritmo de *place learning*, e validar e comparar esses resultados com trabalhos de investigação semelhantes. Como estavam disponíveis *datasets* de apenas 5 utilizadores distintos e respetivos diários, não são em número desejado para a confirmação quer dos parâmetros usados, quer da taxa de acerto de *Places* relativamente aos locais relevantes anotados no diário. Existiam mais utilizadores da aplicação *MOVE*, mas que por falta de diários que servissem de prova dos resultados não foram por isso considerados nos testes e validações realizadas.

No entanto, como o foco deste trabalho não passava pela produção de vistas de mobilidade, quer fossem individuais ou globais, mas sim pelo desenvolvimento de uma estrutura de representação de informação espaçotemporal de mobilidade, os *datasets* disponíveis foram suficientes para a realização da validação pretendida.

6.6. Futuras evoluções

A estrutura de conceitos em si e processos de transformação, está neste momento num ponto em que a sofrer alterações futuras, estas serão mínimas, não sendo expectável modificações significativas na mesma. No entanto, a confirmação só poderá ser feita no momento em que for testada para mais e diferentes fontes de dados, como por exemplo, dados de bilhética dos transportes públicos. O teste da estrutura de conceitos com sensores diferentes dos que foram usados neste trabalho será de interesse para a confirmação da representação produzida e aumentar o potencial aplicacional da mesma.

Como a estrutura de conceitos foi pensada como uma ferramenta de apoio à representação e visualização de informação espaçotemporal, seria interessante ou evoluir a implementação que foi feita e adicionar uma componente própria de visualização de informação de mobilidade ou então integrar a mesma em ferramentas existentes de auxílio à visualização.

O algoritmo de *place learning* como não foi suficientemente testado por falta de mais *datasets* e respetivos diários, necessita por isso de uma validação com um número mais alargado de utilizadores, de forma a confirmar ou ajustar os parâmetros que foram usados (o algoritmo pode sofrer de *overfitting* e responder bem apenas aos utilizadores previamente usados) e desta forma confirmar-se o potencial do modelo probabilístico que foi desenvolvido neste trabalho.

Como não poderia deixar de ser, e em conformidade com a definição de Ecossistema Urbano, seria francamente desejável que tanto a estrutura de conceitos como a implementação desenvolvida pudessem ser usados em ambiente cooperativo com decisores ou planeadores e respetivas equipas multidisciplinares. Por vezes a dificuldade não está na visualização da informação de mobilidade humana, está antes em definir o que se pretende efetivamente visualizar.

Bibliografia

Ahdi, F. *et al.* (2012) *Traffic data collection and anonymous vehicle detection using wireless sensor networks*. Disponível em: http://www.roads.maryland.gov/OPR_Research/MD-12-SP009B4H_Anonymous-Vehicle-Detection-using-Wireless-Sensor-Report.pdf.

Alfeld, L. E. (1995) «Urban dynamics-The first fifty years», *System Dynamics Review*, 11, pp. 199–217. doi: 10.1002/sdr.4260110303.

Andrienko, G. *et al.* (2010) «Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns», *Computer Graphics Forum*, 29, pp. 913–922. doi: 10.1111/j.1467-8659.2009.01664.x.

Andrienko, G. *et al.* (2011) «From movement tracks through events to places: Extracting and characterizing significant places from mobility data», em *VAST 2011 - IEEE Conference on Visual Analytics Science and Technology 2011, Proceedings*, pp. 161–170. doi: 10.1109/VAST.2011.6102454.

Andrienko, G., Andrienko, N., Fuchs, G., *et al.* (2013) «Extracting Semantics of Individual Places from Movement Data by Analyzing Temporal Patterns of Visits», em *Proceedings of The First ACM SIGSPATIAL International Workshop on Computational Models of Place*, pp. 9–16. doi: 10.1145/2534848.2534851.

Andrienko, G. *et al.* (2013) «Scalable analysis of movement data for extracting and exploring significant places», *IEEE Transactions on Visualization and Computer Graphics*, 19, pp. 1078–1094. doi: 10.1109/TVCG.2012.311.

Andrienko, G., Andrienko, N., Bosch, H., *et al.* (2013) «Thematic patterns in georeferenced tweets through space-time visual analytics», *Computing in Science and Engineering*, 15, pp. 72–82. doi: 10.1109/MCSE.2013.70.

Apple Inc. (2015) *iOS 7: compreender os Serviços de localização*. Disponível em: <https://support.apple.com/pt-pt/HT5594>.

Ashbrook, D. e Starner, T. (2002) «Learning significant locations and predicting user movement with GPS», *Proceedings. Sixth International Symposium on Wearable Computers*,. doi: 10.1109/ISWC.2002.1167224.

Authority, G. L. (2013) *THE MAYOR'S VISION FOR CYCLING IN LONDON – An Olympic Legacy for all Londoners, Greater London Authority*. Disponível em: http://www.london.gov.uk/sites/default/files/Cycling_Vision_GLA_template_FINAL.pdf.

Bacon, J. *et al.* (2011) «Using Real-Time Road Traffic Data to Evaluate Congestion», *Dependable and Historic Computing*. doi: 10.1007/978-3-642-24541-1.

Bennett, C., Raab, C. e Regan, P. (2003) «People and Place: Patterns of Individual Identification within Intelligent Transportation Systems», em *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*, pp. 153–175.

Bhattacharya, S. *et al.* (2013) «Gaussian process-based predictive modeling for bus ridership», *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication - UbiComp '13 Adjunct*. New York, New York, USA: ACM Press, pp. 1189–1198. doi: 10.1145/2494091.2497349.

Bloch, M., Fessenden, F. e Carter, S. (2009) *Tracking Taxi Flow Across the City, The New York Times*. Disponível em: <http://www.nytimes.com/interactive/2010/04/02/nyregion/taxi-map.html>.

Brockmann, D., Hufnagel, L. e Geisel, T. (2006) «The scaling laws of human travel.», *Nature*, 439(Ldd), pp. 462–465. doi: 10.1038/nature04292.

Brockmann, D. e Theis, F. (2008) «Money circulation, trackable items, and the emergence of universal human mobility patterns», *Pervasive Computing, IEEE*. IEEE Educational Activities Department, 7(4), pp. 28–35.

Bruner, J. (2012) *American Migration, Forbes*. Disponível em: <http://www.forbes.com/special-report/2011/migration.html>.

Cabalero, L. e Belmonte, N. G. (2013) *World Flights, SenchaLabs*. Disponível em: <http://www.senchalabs.org/philogl/PhiloGL/examples/worldFlights/>.

Chang, J., Sun, E. e Ave, S. C. (2010) «Location : How Users Share and Respond to Location-Based Data on Social Networking Sites», *Artificial Intelligence*, pp. 74–80.

Cheng, Z., Caverlee, J. e Lee, K. (2010) «You Are Where You Tweet : A Content-Based Approach to Geo-locating Twitter Users», *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 759–768. doi: 10.1145/1871437.1871535.

Chon, Y. e Cha, H. (2011) «LifeMap: A Smartphone-Based Context Provider for Location-Based Services», *IEEE Pervasive Computing*. Editado por Z. Rengel. IEEE Computer Society, 10(2), pp. 58–67.

D'Hondt, E., Stevens, M. e Jacobs, A. (2013) «Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring», *Pervasive and Mobile Computing*, 9(5), pp. 681–694. doi: 10.1016/j.pmcj.2012.09.002.

Díaz-Palacios-Sisternes, S., Ayuga, F. e García, A. I. (2014) «A method for detecting and describing land use transformations: An examination of Madrid's southern urban–rural gradient between 1990 and 2006», *Cities*. Elsevier Ltd, 40, pp. 99–110. doi: 10.1016/j.cities.2014.03.010.

Dodge, M. e Kitchin, R. (2006) «Code, vehicles and governmentality: The automatic production of driving spaces», *Maynooth, Ireland: NIRSA Working Papers Series*, 29.

Ert, L. e Steinbach, M. (2003) «Finding Clusters of Different Sizes , Shapes , and Densities in Noisy , High Dimensional Data», *SIAM international conference on data mining*, 47, pp. 47–58.

ESRI (2016) *Metro*. Disponível em: <http://thunderhead.esri.com/readonlyurl/MBTA/MBTA1.html>.

Ester, M. *et al.* (1996) «A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise», *Computer*. Editado por E. Simoudis, J. Han, e U. M. Fayyad. AAAI Press, 1996(6), pp. 226–231. doi: 10.1.1.71.1980.

Ferreira, N., Poco, J. e Vo, H. (2013) «Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips», *Visualization and Computer Graphics, IEEE Transactions on*, 19(12), pp. 2149–2158.

Foell, S. *et al.* (2014) «Catch Me If You Can: Predicting Mobility Patterns of Public Transport Users», *17th International IEEE Conference on Intelligent Transportation Systems*.

Frias-Martinez, V. e Frias-Martinez, E. (2014) «Spectral clustering for sensing urban land use using Twitter activity», *Engineering Applications of Artificial Intelligence*. Elsevier, 35, pp. 237–245. doi: 10.1016/j.engappai.2014.06.019.

Froehlich, J., Neumann, J. e Oliver, N. (2009) «Sensing and predicting the pulse of the city through shared bicycling», em *IJCAI International Joint Conference on Artificial Intelligence*, pp. 1420–1426. doi: 10.1.1.150.4370.

González, M. C., Hidalgo, C. a e Barabási, A.-L. (2008) «Understanding individual human mobility patterns.», *Nature*, 453(7196), pp. 779–82. doi: 10.1038/nature06958.

Google (2015) *Google Location History*. Disponível em: https://support.google.com/accounts/answer/3118687?hl=en&ref_topic=6179443.

Grauwin, S. *et al.* (2014) «Towards a comparative science of cities: using mobile traffic records in New York, London and Hong Kong», *arXiv preprint arXiv:1406.4400*, pp. 1–24.

Hamilton, A. *et al.* (2005) «Urban information model for city planning», *Electronic Journal of Information Technology in Construction*, 10(April), pp. 55–67.

Hasan, S. *et al.* (2013) «Spatiotemporal Patterns of Urban Human Mobility.», *Journal of Statistical Physics*, 151, pp. 304–318. doi: 10.1007/s10955-012-0645-0.

Hightower, J., Consolvo, S. e LaMarca, A. (2005) «Learning and recognizing the places we go», *UbiComp 2005*:

Ubiquitous Computing, pp. 159–176. doi: 10.1007/11551201_10.

Huberman, B. A. and, Romero, D. M. and e Wu, F. (2009) «Social Networks that Matter: Twitter Under the Microscope», *First Monday*, 14(1).

Jansen, B. J. *et al.* (2009) «Twitter power: Tweets as electronic word of mouth», *Journal of the American Society for Information Science and Technology*, 60, pp. 2169–2188. doi: 10.1002/asi.21149.

Jiang, S., Ferreira, J. e Gonzalez, M. C. (2012) «Discovering urban spatial-temporal structure from human activity patterns», *Proceedings of the ACM SIGKDD International Workshop on Urban Computing - UrbComp '12*. New York, New York, USA: ACM Press, p. 95. doi: 10.1145/2346496.2346512.

Kang, J. H. *et al.* (2005) «Extracting places from traces of locations», *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(3), p. 58. doi: 10.1145/1094549.1094558.

Kim, D. H. *et al.* (2009) «Discovering semantically meaningful places from pervasive RF-beacons», em *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, pp. 21–30. doi: 10.1145/1620545.1620549.

Klepeis, N. E. *et al.* (2001) «The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants.», *Journal of exposure analysis and environmental epidemiology*, 11(3), pp. 231–252. doi: 10.1038/sj.jea.7500165.

Koller, D., Daniilidis, K. e Nagel, H. H. (1993) «Model-Based object tracking in monocular sequences of road trafficscenes», *International Journal of Computer Vision*, 10, pp. 257–281. doi: 10.1007/BF01539538.

Krumm, J., Caruana, R. e Counts, S. (2013) «Learning likely locations», *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7899 LNCS, pp. 64–76. doi: 10.1007/978-3-642-38844-6_6.

Lane, N. e Eisenman, S. (2008) «Urban sensing systems: opportunistic or participatory?», em *Proceedings of the 9th workshop on Mobile computing systems and applications*. ACM, pp. 11–16.

Lind, P. G. e Moreira, A. (2015) «Human mobility patterns at the smallest scales», *Communications in Computational Physics*, 18(2), pp. 417–428.

Liu, L., Biderman, A. e Ratti, C. (2009) «Urban Mobility Landscape : Real Time Monitoring of Urban Mobility Patterns», em *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management*, pp. 1–16.

Liu, Y. *et al.* (2014) «Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data», *PLoS ONE*, 9(1). doi: 10.1371/journal.pone.0086026.

MacEachren, A. e Robinson, A. (2011) «Geo-twitter analytics: Applications in crisis management», *Proceedings of the 25th International cartographic conference*, pp. 1–8.

Marmasse, N. e Schmandt, C. (2000) «Location-aware information delivery with comMotion», *Handheld and Ubiquitous Computing*, 1927(2), pp. 157–171. doi: 10.1007/3-540-39959-3_12.

Mathioudakis, M. e Koudas, N. (2010) «Twittermonitor: trend detection over the twitter stream», em *SIGMOD '10 Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pp. 1155–1158. doi: 10.1145/1807167.1807306.

Nassar, A. K., Alan Blackburn, G. e Duncan Whyatt, J. (2014) «Developing the desert: The pace and process of urban growth in Dubai», *Computers, Environment and Urban Systems*. Elsevier Ltd, 45, pp. 50–62. doi: 10.1016/j.compenurbsys.2014.02.005.

National Imagery and Mapping Agency (NIMA) (2000) «Department of Defense World Geodetic System 1984», p. 175. Disponível em: <http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf>.

Peixoto, J. e Moreira, A. (2012) «Dealing with Multiple Source Spatio-temporal Data in Urban Dynamics Analysis», em Murgante, B. *et al.* (eds.) *Computational Science and ITS Applications – ICCSA 2012*. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 450–465.

Peixoto, J. e Moreira, A. (2013) «Human Movement Analysis Using Heterogeneous Data Sources», *International Journal of Agricultural and Environmental Information Systems*, 4(3), pp. 98–117. doi: 10.4018/ijaeis.2013070106.

Perttunen, M. *et al.* (2015) «Urban traffic analysis through multi-modal sensing», *Personal and Ubiquitous*

Computing, 19(3), pp. 709–721. doi: 10.1007/s00779-015-0833-4.

Pinelli, F. *et al.* (2009) «Space and time-dependant bus accessibility: A case study in Rome», *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 346–351. doi: 10.1109/ITSC.2009.5309679.

Planning, C. M. A. for (2014) *CMAP Illinois*. Disponível em: <http://www.cmap.illinois.gov/mobility/explore#/topic/roads/congestion>.

Ramachandra, T. V., Bharath, A. H. e Sowmyashree, M. V. (2014) «Monitoring urbanization and its implications in a mega city from space: Spatiotemporal patterns and its indicators», *Journal of Environmental Management*. doi: 10.1016/j.jenvman.2014.02.015.

Reades, J. *et al.* (2007) «Cellular census: Explorations in urban data collection», *Pervasive Computing, IEEE*, pp. 30–38.

Ribeiro, A. I. J. T. *et al.* (2014) «Studying traffic conditions by analyzing foursquare and instagram data», *Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks - PE-WASUN '14*. New York, New York, USA: ACM Press, pp. 17–24. doi: 10.1145/2653481.2653491.

Sakaki, T., Okazaki, M. e Matsuo, Y. (2010) «Earthquake shakes Twitter users: real-time event detection by social sensors», *Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860.

Santos, J. O. *et al.* (2012) «Semantic Enrichment of Places for the Portuguese Language», *INForum2012*.

Sarwat, M. *et al.* (2014) «LARS *: An Efficient and Scalable Location-Aware Recommender System», *Knowledge and Data Engineering, IEEE Transactions on*, 26(6), pp. 1384–1399.

Sevtsuk, A. e Ratti, C. (2005) «iSPOTS. How wireless technology is changing life on the MIT campus», *Proceedings of the 9th International Conference on Computers in Urban Planning and Urban Management CUPUM 2005*.

Song, C. *et al.* (2010) «Limits of predictability in human mobility.», *Science (New York, N.Y.)*, 327, pp. 1018–1021. doi: 10.1126/science.1177170.

Van der Spek, S. *et al.* (2009) «Sensing Human Activity: GPS Tracking.», *Sensors*, 9(4), pp. 3033–55. doi: 10.3390/s90403033.

Taubenböck, H. *et al.* (2012) «Monitoring urbanization in mega cities from space», *Remote Sensing of Environment*. Elsevier Inc., 117, pp. 162–176. doi: 10.1016/j.rse.2011.09.015.

Thomas, J. e Cook, K. (2005) *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr.

Xia, C. *et al.* (2014) «CityBeat: real-time social media visualization of hyper-local city data», em *Proceedings of the International World Wide Web Conference Committee (IW3C2)*. doi: 10.1145/2567948.2577020.

Yau, N. (2014) *When Do Americans Leave For Work?*, *Flowingdata*. Disponível em: <http://flowingdata.com/2015/02/04/when-do-americans-leave-for-work/>.

Young, S. D., Rivers, C. e Lewis, B. (2014) «Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes», *Preventive Medicine*, 63, pp. 112–115. doi: 10.1016/j.jypmed.2014.01.024.

Yu, H. e Shaw, S. (2004) «Representing and Visualizing Travel Diary Data: A Spatio-temporal GIS Approach», *2004 ESRI International User Conference*, pp. 1–13.

Zheng, Y. *et al.* (2009) «Mining interesting locations and travel sequences from GPS trajectories», *Proceedings of the 18th International Conference on World Wide Web*, (49).

Zhou, C. *et al.* (2007) «Discovering personally meaningful places», *ACM Transactions on Information Systems*, 25(3). doi: 10.1145/1247715.1247718.